A / REISSUE

JC544 U.S. PTO 07/07/98

PTO/SB/17 (12-97)
Approved for use through 9/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL

Note: Effective October 1, 1997.
Patent fees are subject to annual revision.

TOTAL AMOUNT OF PAYMENT ($) 518.00

Complete if Known	
Application Number	
Filing Date	
First Named Inventor	Chong, L.
Group Art Unit	
Examiner Name	
Attorney Docket Number	TLC-RE

METHOD OF PAYMENT (check one)

1. ☐ The Commissioner is hereby authorized to charge any deficiency indicated fees and credit any over payments to:

Deposit Account Number: 15-0699
Deposit Account Name: Ostrager Chong et al

☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17 ☐ Charge the Issue Fee Set in 37 CFR 1.18 at the Mailing of the Notice of Allowance

2. ☒ Payment Enclosed: $518.00
☒ Check ☐ Money Order ☐ Other

FEE CALCULATION

1. FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
101	790	201	395	Utility filing fee	
106	330	206	165	Design filing fee	
107	540	207	270	Plant filing fee	
108	790	208	395	Reissue filing fee	395
114	150	214	75	Provisional filing fee	

SUBTOTAL (1) ($) 395

2. CLAIMS

		Extra	Fee from below	Fee Paid
Total Claims	20 -20 =	0	X =	
Independent Claims	4 -3 =	3	X 41 =	123
Multiple Dependent Claims			X =	

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
103	22	203	11	Claims in excess of 20
102	82	202	41	Independent claims in excess of 3
104	270	204	135	Multiple dependent claim
109	82	209	41	Reissue independent claims over original patent
110	22	210	11	Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) 123

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
105	130	205	65	Surcharge - late filing fee or oath	
127	50	227	25	Surcharge - late provisional filing fee or cover sheet.	
139	130	139	130	Non-English specification	
147	2,520	147	2,520	For filing a request for reexamination	
112	920*	112	920*	Requesting publication of SIR prior to Examiner action	
113	1,840*	113	1,840*	Requesting publication of SIR after Examiner action	
115	110	215	55	Extension for reply within first month	
116	400	216	200	Extension for reply within second month	
117	950	217	475	Extension for reply within third month	
118	1,510	218	755	Extension for reply within fourth month	
128	2,060	228	1,030	Extension for reply within fifth month	
119	310	219	155	Notice of Appeal	
120	310	220	155	Filing a brief in support of an appeal	
121	270	221	135	Request for oral hearing	
138	1,510	138	1,510	Petition to institute a public use proceeding	
140	110	240	55	Petition to revive - unavoidable	
141	1,320	241	660	Petition to revive - unintentional	
142	1,320	242	660	Utility issue fee (or reissue)	
143	450	243	225	Design issue fee	
144	670	244	335	Plant issue fee	
122	130	122	130	Petitions to the Commissioner	
123	50	123	50	Petitions related to provisional applications	
126	240	126	240	Submission of Information Disclosure Stmt	
581	40	581	40	Recording each patent assignment per property (times number of properties)	
146	790	246	395	Filing a submission after final rejection (37 CFR 1.129(a))	
149	790	249	395	For each additional invention to be examined (37 CFR 1.129(b))	

Other fee (specify) _____

Other fee (specify) _____

* Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($)

SUBMITTED BY				Complete (if applicable)	
Typed or Printed Name	Leighton K. Chong			Reg. Number	27,621
Signature	Leighton K. Chong	Date	7/7/98	Deposit Account User ID	LKC

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/96 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

CERTIFICATE UNDER 37 CFR 3.73(b)

Applicant: Leighton K. CHONG, Christine K. KAMPRATH

Application No.: U.S. Patent 5,535,120 Filed: Issued July 9, 1998
from Appln. 08/487,450, filed June 7, 1995

Entitled: MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ID DATA TO SELECT DICTIONARIES

TRANS-LINK INT'L CORP., a corporation of the State of Hawaii,
(Name of Assignee) (Type of Assignee, e.g , corporation, partnership, university, government agency, etc.)

certifies that it is:

1. [XX] the assignee of the entire right, title, and interest; or

2. [ ] an assignee of an undivided part interest

in the patent application identified above by virtue of either:

A. [XX] An assignment from the inventor(s) of the patent application identified above. The assignment was recorded in the Patent and Trademark Office at Reel 7912, Frame 0414, or for which a copy thereof is attached.

OR

B. [ ] A chain of title from the inventor(s), of the patent application identified above, to the current assignee as shown below:

1. From:_____ To:_____
The document was recorded in the Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

2. From:_____ To:_____
The document was recorded in the Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

3. From:_____ To:_____
The document was recorded in the Patent and Trademark Office at
Reel _____, Frame _____, or for which a copy thereof is attached.

[ ] Additional documents in the chain of title are listed on a supplemental sheet.

[XX] Copies of assignments or other documents in the chain of title are attached.

The undersigned (whose title is supplied below) is empowered to sign this certificate on behalf of the assignee.

June 22, 1998
Date

Leighton K Chong
Signature

Leighton K. Chong, President
Typed or printed name

TRANS-LINK INTERNATIONAL CORP.
Title

MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING [USER ID] CONTROL DATA TO SELECT LANGUAGES AND DICTIONARIES

SPECIFICATION

Technical Field

This invention relates to a system for automatic (machine) translation of text and, more particularly, to a telecommunications-based system for automatically translating and sending text from a sender to a recipient in another language.

Background Art

After several decades of development, the field of automatic (machine) translation of text from a source language to a target language with a minimum of human intervention has developed to a rudimentary level where machine translation systems with limited vocabularies or limited language environments can produce a basic level of acceptably translated text. Some current systems can produce translations for unconstrained input in a selected language pair, i.e., from a chosen source language to a chosen target language, that is perhaps 50% acceptable to a native writer in the target language (using an arbitrary scale measure). When the translation system is constrained to a particular vocabulary or syntax style of a limited area of application (referred to as a "sublanguage"), the results that can now be achieved may approach a

level 90% acceptable to a native writer. The wide difference in results is attributable to the difficulty of producing acceptable translations when the system must encompass a wide variability in vocabulary use, syntax, and expression, as compared to the limited vocabularies and translation equivalents of a chosen sublanguage.

One example of a machine translation system limited to a specific sublanguage application is the TAUM-METEO system developed by the University of Montreal for translating weather reports issued by the Canadian Environment Department from English into French. TAUM-METEO uses the transfer method of translation, which consists basically of the three steps of: (1) analyzing the sequence and morphological forms of input words of the source language and determining their phrase and sentence structure, (2) transferring (directly translating) the input text into sentences of equivalent words of the target language using dictionary look-up and a developed set of transfer rules for word and/or phrase selections; then (3) synthesizing an acceptable output text in the target language using developed rules for target language syntax and grammar. TAUM-METEO was designed to operate for the English-French language pair in the narrow sublanguage of meteorology (1,500 dictionary entries, including several hundred place names; input texts containing no tensed verbs). It therefore can obtain high levels of translation accuracy of 80% to 90% by avoiding the need for any significant level of morphological analysis of input words, by analyzing input texts for domain-specific word markers which narrow the range of choices for output word selection and syntax structure, and by using ad hoc transfer rules for output word and phrase selections.

Another example of a sublanguage translation system is the

METAL system developed by the Linguistics Research Center at the University of Texas at Austin for large-volume translations from German into English of texts in the field of telecommunications. The METAL system also uses the transfer method, but adds a fourth step called "integration" between the analysis and transfer steps. The integration step attempts to reduce the variability of output word selection and syntax by performing tests on the constituent words of the input text strings and constraining their application based upon developed grammar and phrase structure rules. Transfer dictionaries typically consist of roughly 10,000 word pairs. In terms of translation quality, the METAL system is reported to have achieved between 45% and 85% correct translations.

A strategy competing with the transfer approach is the "interlingua" approach which attempts to decompile input texts of a source language into an intermediate language which represents their "meaning" or semantic content, and then convert the semantic structures into equivalent output sentences of a target language by using a knowledge base of contextual, lexical, and syntactic rules. Historically, transfer systems lacking a comprehensive knowledge base and limited to translation of sentences in isolation have had the central problem of obtaining accurate word and phrase selections in the face of ambiguities presented by homonyms, polysemic phrases, and anaphoric references. The interlingua approach is favored because its representation of text meaning within a context larger than single sentences can, in theory, greatly reduce ambiguity in the analysis of input texts. Also, once the input text has been decompiled into a semantic structure, it can theoretically be translated into multiple target languages using the linguistic and semantic rules developed for each target language. In practice,

however, the interlingua approach has proven difficult to implement because it requires the development of a universal symbolic language for representing "meaning" and comprehensive knowledge bases for making the conversions from source to intermediate and then to target languages. Examples of interlingua systems include the Distributed Translation Language (DLT) undertaken in Utrecht, the Netherlands, and the Knowledge-Based Machine Translation (KBMT) system of the Center for Machine Translation at Carnegie-Mellon University.

Other machine translation systems have been developed or are under development using modifications or hybrids of the transfer and interlingua approaches. For example, some systems use human pre-editing and/or post-editing to reduce text ambiguity and improve the correctness of word and phrase selections. Other systems attempt to combine a basic transfer approach with knowledge bases and artificial intelligence techniques for machine editing and enhancement. Another approach is to combine decompilation to a syntactically-based intermediate structure with transfer to equivalent output phrases and sentences. For a further discussion of current developments in machine translation, reference is made to Machine Translation, Theoretical and Methodological Issues, edited by Sergei Nirenberg, published by Cambridge University Press, 1987, and "Proceedings of The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language", published by Linguistics Research Center, University of Texas at Austin, June 1990.

It is expected that machine translation (MT) systems will develop in time to provide higher levels of translation accuracy and utility. However, current MT techniques using a basic transfer

approach can produce acceptable translation accuracy in a selected sublanguage, yet they are not in widespread use. One reason for the limited use of MT systems is that most current systems are designed for a single, specific application, environment and language pair context. The requirements of that context motivate the design and development of the grammar, dictionary structure, and parsing algorithms. Thus, the utility of the system becomes confined to that particular context. This approach greatly limits the range of applications and the audience of users which can be productively served by such application- and language-specific MT systems.

Summary of Invention

It is therefore a principal object of the present invention to provide a system which can perform machine translation among a plurality of source languages, target languages, and sublanguages, and automatically send the translated text via telecommunications links to one or more recipients in different languages and/or in different locations. The system should be capable of providing acceptable levels of translation accuracy and be readily upgradable to higher levels of accuracy and utility. It is a further object that such a system be capable of operation with a minimum of human intervention, yet have interactive utilities for obtaining and adding new word entries to its dictionary database. It is also desired that such a system be capable of building and organizing a large-scale dictionary database containing core language dictionaries, plural sublanguage dictionaries, and individual user dictionaries in a manner which cumulates and evolves over time.

In accordance with a principal aspect of the present invention, a machine translation and telecommunications system comprises:

(a) a machine translation module which is capable of performing machine translation from input text of a source language to output text of a target language;

(b) a receiving interface for receiving input via a first telecommunications link, said input including an input text to be translated accompanied by a control portion having at least a first predefined field therein for designating an address of a recipient to which translated output text is to be sent;

(c) a recognition module coupled to said receiving interface for electronically scanning the control portion and recognizing the address of the recipient designated in the first predefined field of the control portion; and

(d) an output module including a sending interface for sending translated output text generated by said machine translation module to the address of the recipient recognized by said recognition module via a second telecommunications link.

In a more specific aspect of the invention relating to [sublanguage] source/target language selection, a machine translation system comprises:

(a) a receiving interface for receiving an input text in a source language accompanied by [and a sublanguage] a control input including at least a first predefined field containing an address of a receipient to receive output text in a target language and a second predefined field containing a source/target language control input indicative of a selected [sublanguage] source/target language pair for translation applicable to the input text from among a plurality

of possible [sublanguages] source/target language pairs;

(b) a machine translation module capable of performing machine translation of an input text in a source language to an output text in a target language using a dictionary database containing entries for words of the target language corresponding to words of the source language;

(c) a dictionary database containing a plurality of source/target language dictionaries, each [including a core language dictionary] containing entries for generic words of [the] a source and target language[s] pair [, and a plurality of sublanguage dictionaries each containing entries for specialized words of a respective sublanguage];

(d) a dictionary control module responsive to the [sublanguage] source/target language control input for selecting a [sublanguage] source/target language dictionary of the dictionary database which is applicable to the input text, and for causing the machine translation module to use the selected [sublanguage] source/target language dictionary in performing translation of the input text; and

(e) an output module responsive to the address of the control input for outputting translated text in the target language generated by the machine translation module and automatically routing it to be sent to the recipient's address.

In another aspect of the present invention, [the] a sublanguage control input causes a selected sublanguage dictionary deemed especially applicable to the input text to be used in order to perform more accurate translation of the input text. The dictionary database includes core and sublanguage dictionaries for a plurality of source/target languages and sublanguages. The machine translation

system with this multiple core languages and sublanguages capability is employed in a telecommunications system which automatically translates and transmits text from [a] one or more senders to one or more recipients in [other] different languages. A cover page or header accompanying the input text is used to designate the selected source/target languages, the applicable sublanguages, and the address(es) -- electronic, fax, or mail -- of the recipient(s).

In a preferred embodiment, the receiving interface receives input text as electronic (machine-readable) text over a communications line, or as page image data via a fax/modem board or page scanner. The receiving interface is operated in a computer server along with a recognition module for converting any page image data to electronic text. The recognition module scans and recognizes designations of the cover page or header accompanying the input text for determining the selections of the source/target languages and sublanguage(s) applicable to the input text. In the case of electronic text, the cover page and the input text may be introduced by means of a disk file, by downloading an electronic file, or by online user-system interaction. An optional online interaction mode can prompt the user for information concerning the user's identity, sublanguage preferences, and/or a particular input text to be translated in order to facilitate generation of a suitable cover page. Inferencing algorithms may be used to assess the user and cover page information and determine the applicable sublanguage dictionary(ies).

The output module may have a page formatting program for composing the translated output text into a desired page format appropriate to a particular recipient or target language. It may

also have a footnoting function for providing footnotes of ambiguous phrases of the input text in their original source language and/or with alternate translations in the target language. The output module includes a sending interface coupled to a fax/modem board for facsimile transmission, or a printer for printing output pages, or a telecommunications interface for sending output electronic text to a recipient's electronic address. The modularity of the receiving interface, dictionary database, dictionary control module, and output module from the machine translation module assures that, as machine translation improvements are developed, the machine translation module may be upgraded or replaced without rendering the other portions of the system dysfunctional or obsolete.

As another aspect of the invention related to a machine dictionary database, a machine translation system comprises:

(a) a machine translation module capable of performing machine translation of input text in a source language to output text in a target language using a dictionary database containing entries for words of the target language corresponding to words of the source language;

(b) a dictionary database including a core language dictionary containing entries for generic words of the source/target languages, a plurality of sublanguage dictionaries each containing entries for specialized words of a respective sublanguage used by a group of users, and a plurality of user dictionaries each containing entries for individualized words of a respective user; and

(c) a dictionary control module responsive to control inputs to the machine translation system for causing the machine translation module to use the core language dictionary, any applicable sublanguage dictionary, and any applicable user dictionary

for performing translation of an input text attributed to a user of the system.

In this aspect of the invention, a large-scale dictionary database is maintained which has dictionaries containing word entries specified linguistically at different hierarchical levels of usage. At the lowest (user) level, a particular user can enter temporary or "scratch" word entries into a respective user dictionary. The machine translation system uses the particular user's dictionary to perform machine translation of text which may contain idiosyncratic or new words or phrases particularly used by that user. The dictionary control module includes dictionary maintenance utilities which allow such scratch entries to be entered by users into their user dictionaries, and which assist a dictionary maintenance operator (DMO) to review the scratch entries so that they can be confirmed as valid dictionary entries for machine translation. The dictionary maintenance utilities include automated programmed procedures for assessing whether particular word entries appearing in the user dictionaries can be moved into a higher-level sublanguage dictionary for a given domain or group of users. In a similar manner, word entries that appear in common in the sublanguage dictionaries of a wide range of domains or user groups can be moved as generic word entries into the core language dictionary.

Other objects, features, and advantages of the present invention will become apparent from the following detailed description of the preferred embodiments of the invention, as considered with reference to the following drawings:

Brief Description of Drawings

Fig. 1 is a schematic diagram of a machine translation and telecommunications system in accordance with the invention.

Fig. 1A is a schematic diagram of a computer server which includes a receiving interface, recognition module, and dictionary control module, and is coupled to a machine translation module and an output module.

Fig. 1B is a schematic diagram of a machine translation module which includes a translation processing module and a dictionary database, and its linkage to the computer server and the output module.

Fig. 1C is a schematic diagram of the output module, including a page formatting module and a sending interface.

Fig. 2 is an illustration of a cover page for designating core language pair, sublanguage(s), and recipient information, and accompanying text pages.

Fig. 3 is an illustration of input ideographic text and output English text as performed by the machine translation system using page formatting functions.

Fig. 4 is a schematic diagram of the dictionary control module, comprising a dictionary selection submodule and a dictionary maintenance submodule, wherein the latter includes an (interactive) user maintenance module and a dictionary maintenance module as

overseen by a dictionary maintenance operator (DMO).

Fig. 5 is a schematic representation of an interactive input editor for interactions with users of the system.

Fig. 6 is a schematic diagram illustrating dictionary maintenance utilities for collapsing and promotion of entries from subordinate to superordinate dictionaries.

Fig. 7A illustrates, as a function of the dictionary maintenance utilities, the creation of scratch word entry from an identical word entry.

Fig. 7B illustrates the use of utilities with an interactive input editor to scan various levels of the dictionary hierarchy for word entries on which to base scratch word entries.

Fig. 7C illustrates a typical content of an identical word entry from which a scratch word entry is created.

Fig. 7D illustrates the creation of a "copy-cat" word entry from a synonymous word entry.

Detailed Description of Preferred Embodiments

Referring to Fig. 1, a preferred form of the machine translation and telecommunications system in accordance with the present invention comprises a computer server 10, a machine translation module 20, and an output module 30. (These and further-

described components of the system will be denoted with capital letters for clarity of reference.) The Computer Server 10 receives electronic text input accompanied by a cover page or header from any of a plurality of input sources, designated generally as a telecommunications link A. The Computer Server 10 has a function for recognizing control data in the cover page or header designating core language and sublanguage selections applicable to the input text to be translated. It also recognizes output addresses and page formatting data to be used by the Output Module 30 for transmitting the translated text to the designated recipient(s) via any of a plurality of output devices, designated generally as a telecommunications link B. Due to the modularity of the system, the Machine Translation Module 20 may be updated by operator maintenance or upgraded or replaced without rendering the other functions of the system dysfunctional or obsolete.

The Machine Translation Module 20 is capable of performing machine translation from input text in a source language to output text in a target language. In the examples of a machine translation (MT) system described herein, reference is made generally to an MT system of the transfer type which relies upon the use of a machine-readable dictionary for lookup of source/target word entries. The principles of the present invention may also be applied to an MT system of the interlingua type. Transfer-type MT systems are more widely accepted for near-term usage than interlingua systems, and they rely more heavily on linguistic knowledge incorporated into machine dictionaries designed for source/target language pairs. The operation of transfer-type MT systems is well understood by those skilled in the machine translation field, and is not described in detail herein.

Input Data Reception and Extraction

Fig. 1A shows the Computer Server 10 having a Receiving Interface 11 linked to the telecommunications link A, a Recognition Module 12, and a Dictionary Control Module 13. The Receiving Interface 11 may include an interactive mode program (to be described further herein) whereby a user can provide cover page or header designations, update or create User ID files pertinent to translation parameters associated with that user's communications, or create specialized user dictionary entries during interactive text entry sessions. The Recognition Module 12 includes a character recognition (often referred to as "OCR") program which recognizes and converts page image data into machine-readable text, and which recognizes cover page designations or user designations referencing cover page data stored in the User ID files. The Dictionary Control Module 13 includes a Dictionary Selection Module, which assesses the control data it receives from the Recognition Module 12 and designates the appropriate core language and sublanguage dictionary(ies) to be used by the Machine Translation Module 20. It also includes a Dictionary Maintenance Module, which allows a dictionary maintenance operator (DMO) to create and update dictionary entries in the Dictionary Database 22.

Using the control data from a cover page or header accompanying the input text, the Computer Server 10 allows the system to automatically recognize a sender's designations of the source language of the input text, the target language(s) of the output text, any particular sublanguage(s) used in a specialized domain, user group, or correspondence type, any preferred page format for the

output text, and the address(es) of one or more recipients to whom the translated text is to be sent. Thus, the system can automatically access designated ones of the plurality of core and sublanguage dictionaries maintained in the Dictionary Database 22 for different source/target languages and sublanguages, and can format and transmit the translated text to recipient(s) in respective target language(s) via telecommunications link B, without the need for any substantial human intervention.

The Computer Server 10 interfaces with a plurality of receiving devices. For example, input data can be received as a facsimile transmission via a fax/modem board plugged into the I/O bus for the server system. Such fax/modem boards are widely available and their operation in a server system is well understood by those skilled in this field. Input may also be received from a conventional facsimile machine coupled to a telephone line which prints facsimile pages converted from signals transmitted on the telephone line. A conventional page scanner with a sheet feeder can be used to scan in facsimile or printed pages as page image data for input to the Computer Server. The page image data is then converted to machine-readable form by the OCR program. Input may also be received through a telecommunications program or network interface as electronic text or text files (such as ASCII text), in which case conversion by the OCR program is not required.

The OCR program is preferably resident as an application program in the Computer Server 10 along with the interface programs for handling the reception of input data. OCR programs are widely available, and their operation is well known in this field. For example, an OCR program for scanning and recognizing Japanese kana

and ideographic characters is offered by Catena Corp., Tokyo, Japan. An example of an OCR program for alphanumeric characters is WordScanTM offered by Calera Recognition Systems, Santa Clara, California. The Computer Server 10 is preferably a high-speed, multi-tasking PC computer or workstation.

Referring to Fig. 2, the Computer Server 10 receives input data which is divided into two parts: a cover page or header 50 and input text 60. In the example shown, a cover page is used in conjunction with other pages of input text in a page-oriented system. In the case of transmission of an electronic text file or a text message, a preceeding header or identifier for the communication is used. The cover page 50 has a number of fields for designating selections of source/target language(s), sublanguage(s), page format, and recipient(s) for the text. The cover page 50 is organized with data fields in a predefined format which is readily recognized by the Recognition Module 12 of the Computer Server 10 so that the control data in the predefined fields can be readily recognized.

For example, the cover page 50 may be laid out and formatted with field boundaries and markings on the printed page which can be optically scanned with a high level of reliability. Line dividers 51 and large type-size headers 52 may be used to mark the sender, source/target language(s), sublanguage (communication type or subject matter), page format, and recipient address fields. Boxes 53, which can be marked or blackened in, allow the designated selections to be determined without error. The names of the sender and recipients, their respective companies, addresses, and telephone and/or facsimile transmission numbers are determined by character recognition once the respective fields 51, 52 have been

distinguished. Any page length of input text 60 can follow the cover page 50. Alternatively, information ordinarily supplied by a cover page or header may be stored in the User ID files and supplied automatically as a memorized script in response to user selection.

It is the task of the Recognition Module 12 to extract data pertinent to dictionary selection from the fields of the cover page or header. In batch mode this data is predetermined -- it is either filled into the cover page fields by the user with each specific translation transaction, or it can be supplied by a reference to the User Identification (ID) files resident in the Recognition Module 12.

In the Interactive Mode for specifying the cover page or header through the Receiving Interface 11, the user may first be presented with predetermined sets of fill-in data and then prompted for alternative values, or provided with a variety of alternatives from which to choose, based upon a variety of sets of data already stored in the User ID files, or based upon inferences drawn from the data as it is entered by the user. For example, a User A may specify Recipient Z by name only, and then be presented with additional data, such as Recipient Z's address, title, or affiliation, already stored in the User ID files for verification or correction. Alternatively, Recipient Z may never have been addressed by User A in the past but may be a user categorized in Domain L, which is a domain of which User A is also a member, thus triggering the inference that the sublanguage dictionary of Domain L may be presented to User A as an option for use.

The user may be prompted in Interactive Mode to verify or choose among field values which may aid in selecting one or more

sublanguage dictionaries for a given translation, including correspondence types, subject domains, social indicators, etc. By automating the filling-in of cover page information, the system maximizes the flexibility and accessibility of the system's capabilities for the user while controlling and monitoring the completeness and cohesiveness of the data supplied.

The cover page may designate a plurality of recipients in a plurality of address locations and target languages, each of which may have particular formatting requirements for the output. For automated assistance, each prospective recipient can be referenced by an identifying code indexed to data stored in the User ID files. For example, a travel agent may have a regular set of clients in a variety of locations and languages, with access to a variety of communication modes, to whom he or she regularly sends advertising material. One client may require Japanese translation formatted as "right-to-left" vertical lines of ideographic characters, to be printed and sent as ordinary mail. Another may require faxed translation into German. Still another may have E-mail capability and require a printed copy as well. All these combinations of addressees and requirements can be predefined and stored in the User ID files. The sets of data to be supplied to the cover page fields for each of these addressees may be indexed to mnemonic codes, such as the addressee's alphabetic name, which are supplied by the user, so that they can be retrieved from the User ID files by the Recognition Module.

The User ID files may be established at the time of subscription by a user to a machine translation service, and updated from time to time thereafter. Using the Interactive Mode, the user

may be prompted to supply his or her name, sex, title, company, address, group affiliations, source language, etc., as well as data relevant to prospective recipients or groups of recipients to be stored in the User ID files for filling in cover pages automatically. Sublanguage selections appropriate to the user may be identified or queried by comparing the requirements of the user with those of other users subscribing to the service.

The user may also be prompted to provide samples of typical texts expected to be submitted for translation, as well as individualized or key words for a thesaurus of terms relevant to that user and/or chosen sublanguages. Automatic utilities may be employed to make inferences about the choice of sublanguage dictionaries most appropriate to the new user, based in part on a thesaurus of relevant terms supplied by individual users and groups of users in the same subject domain. The recognition of the user's membership in one or more established groups or subject domains is an important part of dictionary selection and maintenance.

At the time of each translation transaction, the user may be prompted by the system to tailor the cover page to the specific translation transaction about to be initiated. The system may ask the user to confirm a default cover page configuration, to select or modify previously established cover page configurations, or to fill in a scratch cover page which may be blank or partially filled in with data from the User ID files.

Machine Translation Using Sublanguage Dictionaries

As shown in Fig. 1B, the Machine Translation Module 20 is comprised of a Translation Processing Module 21, and a Dictionary Database 22 containing dictionaries for a number of core source/target languages I, II, III, IV, etc., each of which may contain a plurality of domain, subdomain, and user dictionaries. The Translation Processing Module 21 may be a conventional transfer-type system, such as the ECS Natural Language Processing System (hereinafter "ECS/MT system") offered by Executive Communication Systems, Inc., Provo, Utah. The selection indices for the core language and sublanguage dictionaries provided by the Dictionary Selection Module determine which dictionaries in the Dictionary Database are used. The selected dictionaries may be compiled together as one operating dictionary, or prioritized and arranged hierarchically in the system's RAM memory.

The Machine Translation Module is shown as a separate module which receives system data designating the core language and sublanguage(s), if any, to be used and the input text from the Computer Server 10 via the Dictionary Control Module 13. In this manner, the machine translation functions are kept separate from the receiving interface, recognition, User ID files, dictionary selection, dictionary maintenance, and other functions of the Computer Server 10, so that they can be easily upgraded and/or replaced with enhanced programs without disruption to the remainder of the system. The Computer Server 10 acts as a control unit for the Machine Translation Module 20 by exploiting the functions of the Dictionary Control Module 13 for selecting the core language and sublanguage(s) to be used in accordance with the system control data extracted by the Recognition Module 12.

The Computer Server, Machine Translation Module, and Output Module may all reside together on the same workstation. A current target for machine translation systems is a speed of about 20,000 to 30,000 words/hour. A workstation using currently available transfer-type translation programs can attain this range with a processor speed of about 50 to 100 MIPs (million instructions per seconds). Substantial savings on disk access times can be obtained by providing a RAM capacity sufficient to hold all selected core and sublanguage dictionaries in internal memory. For a typical core dictionary size of 60K entries (100 bytes each) for each of the source, transfer, and target lexicons, plus three sublanguage dictionaries of 5K x 3 entries each, as well as system program and operations files, a RAM capacity of the order of 48 MB of internal memory or more is desireable.

Alternatively, the system may be implemented with separate processing units. For example, the Computer Server and Output Module may be implemented as a telecommunications workstation, while the Machine Translation Module may be implemented via a RISC processor, parallel processors, or a supercomputer for high-speed batch processing of multiple source/target language, sublanguages, and output formats.

Machine translation is generally performed by passing each sentence of the text to be translated through a series of stages. Typically, these stages include: (a) source text dictionary lookup and morphological analysis; (b) identification of homographs; (c) identification of compound nouns; (d) processing of prepositions; (e) identification of nouns and verb phrases; (f) subject-predicate identification; (g) syntactic ambiguity identification; (h)

processing of idioms; (i) mapping of source structures onto target structures; (j) synthesis and morphological processing of target text; and (k) rearrangement of words and phrases in target text.

As an example, the ECS/MT System is a transfer-type system based on Lexical Functional Grammar theory and constructs. An input sentence is parsed word by word in left-to-right fashion. Each word is searched by lookup in the source dictionary to determine its morphological, lexical, and syntactic attributes. In the ECS implementation of Lexical Functional Grammar, the indexed attributes of words are used to call analysis routines or invoke grammar rules which enable recognition of the word's place and function within a phrase component of the sentence. Decisions based upon the analysis rules and analysis process assist in disambiguating the lexical meaning and phrase structure of the input sentence in the source language.

The result of parsing in the analysis phase is an intermediary graph or table representing the source-language phrase structure of the sentence, mapped to a directed acyclic graph displaying the grammatical function of words within the sentence and their lexical attributes. This Lexical Functional Grammar representation is largely language independent. During the transfer phase, the functional structure representation of the source-language sentence is transferred by lexical and syntactic transfer rules into an equivalent target-language representation of functional structure and lexical attributes. This target-language representation is then synthesized into an output sentence using the lexical data and grammar rules provided by the target language dictionary. A core language dictionary, including source and target word entries,

bilingual transfer entries, and morphological, syntactic, and lexical rules for both source and target languages, is required for each language pair.

In the present invention, the Translation Processing Module 21 also uses a selected sublanguage dictionary containing specialized word entries and grammar rules specific to a sublanguage that is particularly applicable to the input text. Each sublanguage dictionary set up in the Dictionary Database 22 is chosen to have a manageable size, predictable modes of expression and syntactic structures, and a well-understood context for disambiguation of homonyms, polysemic phrases, and specialized references.

In the machine translation field, the term "sublanguage" usually refers to a recognized domain having a defined set of terms and patterns of language usage that characterize that domain. In the present invention, "sublanguage" is used more loosely to refer to any set of terms and patterns of usage attributed to a field of usage, group of users, or even an individual user. That is, a "sublanguage dictionary" is set up on a fluid or *ad hoc* basis whenever a preferred set of terms and usages is identified.

As illustrated in Fig. 2, for example, designated sublanguages might include correspondence types, such as business letters, legal/technical analysis, technical writing, financial/market reports, or general writing. Business correspondence typically employs only a few pages, a limited vocabulary (on the order of 6000 words), and a limited set of syntactic structures (often restricted to declarative sentences).

Designated sublanguages may also encompass specific fields, e.g., technical fields such as physics, chemistry, electronics, military, etc., or commercial fields such as travel and tourism, real estate, finance, shipping, insurance, etc., or groups of users such as associations, corporations, departments, or simply persons in regular communication with each other.

Sublanguage dictionaries may also be set up corresponding to socially-determined usages or particular contexts in which certain communications take place. For example, in some languages, such as Japanese, certain words, forms of address, and even whole expressions are determined by the relative age, sex, position, grouping (internal/external), or environment of the speaker and the person being addressed. Such particular terms and usages can be set up as distinct sublanguage dictionaries that are accessed according to factors identified in the cover page or header for a communication, e.g., status or sex-indicative titles of the sender and recipient, positions in their respective companies, locations of the sender and recipient, etc.

Setting up sublanguage dictionaries can be implemented with dictionary-building tools currently used in machine translation systems. For example, the ECS/MT system provides a set of tools to develop a core dictionary for a chosen language pair and a technical dictionary for a chosen sublanguage. A Rule Editor tool allows a linguist to create and modify morphological rules, phrase structure rules, and transfer rules for the sublanguage. A Dictionary Maintenance Utility allows creation and modification of lexical entries, including source entries, target entries, and source-to-target transfer entries in the dictionary. A Translation Module

performs table-driven translation using linguistic tables, analysis rules, transfer rules, and semantic preference entries that have been compiled into the dictionary. A Morphology Module applies rules to analyze morphologically complex words to determine uninflected forms for dictionary lookup of source lexical items and to generate morphologically complex words in the target language. A Semantic Preference Component provides for the identification of semantic relations, the assignment of semantic attributes to lexical items, and the accessibility and matching of these attributes for lexical disambiguation and selection of preferred translations.

Dictionary Organization and Selection

In the present invention, core language dictionaries and a plurality of ad hoc sublanguage dictionaries (including both lexical entries and grammar rules for morphological and syntactic analysis and generation) are maintained in the system's Dictionary Database 22. The core language dictionaries are developed and maintained according to linguistic methods and tools commonly used in the machine translation field. In the present invention, sublanguage dictionaries are set up for any identified commercial or technical fields, application domains, groups of users, and even individual users. No particular effort is made to rigorously identify sublanguage boundaries or general sublanguage patterns. Instead, a sublanguage dictionary is set up or updated anytime the vocabulary or syntactic preferences of a user or group of users can be identified. Individual user or lower level dictionaries may be combined or integrated into master (higher-level) sublanguage dictionaries for any field, application domain, or user group when more general

sublanguage vocabulary or syntactic preferences are identified.

The Dictionary Database embodies at once a hierarchical structure of nested dictionaries, arranged in order of generality of usage and exploiting inheritance of linguistic attributes within entries, and a relational structure, whereby the various dictionaries and the particular entries within them inform the establishment of subject domains and the sublanguage dictionaries pertinent to them. The Dictionary Database includes a core language dictionary which contains entries for words in their most general usages, as well as a set of grammar rules for analyzing and generating their morphological and syntactic structures. In the transfer approach, the core dictionary must contain three parallel entries for each term to be translated, i.e., two monolingual entries, one for the source language and one for the target language, containing information about the morphological, syntactic, and semantic characteristics of the word in relation to its own language, and a bilingual transfer entry specifying details required to translate the source word into the target, including information on whatever structural changes must be made during the translation process. The monolingual entries may be usable in a monolingual dictionary for another source/target language pair, however, the bilingual (transfer) entries are specific to the language pair involved.

As illustrated in Fig. 1B, the Dictionary Database of the present invention allows for a multiplicity of levels of nested sublanguage dictionaries along with the core language dictionary. At the lowest level, user dictionaries may exist for individual users. The user dictionaries are nested within higher-level "subdomain" or "master" dictionaries Sub1, Sub2, etc. The subdomain dictionaries

contain more general word entries and grammar/linguistic rules that are common to the users grouped within or cross-referenced to that subdomain. The subdomain dictionaries are nested within higher-level "domain" dictionaries Dom1, Dom2, etc. The domain dictionaries contain even more general word entries and grammar/linguistic rules that are common to the subdomains grouped within or cross-referenced to that domain. At the highest level, the domain dictionaries are nested within the core dictionary which contains common words and rules that are generic to most or all of the included domains. The sublanguage dictionary entries have the same general structure as the core dictionary entries. Thus, in the transfer approach, the two monolingual and third transfer entries for each input word must be available in the sublanguage dictionary.

This hierarchical organization of dictionaries provides for minimum dictionary lookup time in the translation of a sublanguage-specific text, because it directs the lookup first to the user dictionaries, then to the sublanguage dictionaries. When the lower level dictionaries are searched first, a more accurate, efficient, and idiomatic translation is likely to be obtained than the broad-based and more general core dictionary could provide. If a lower level dictionary cannot analyze and resolve an input item, the Dictionary Control Module 13 then accesses a next level dictionary and finally the core dictionary upon failure of translation at a more specific level. The dictionaries are selected for domain specificity by the Dictionary Selection Module, but the progression of access is inherent in the nested structure.

Another aspect of the Dictionary Database lies in its treatment of linguistic information. The generativity of human

grammar relies on the capability of organizing linguistic data into types, such as Noun, Verb, Adjective, etc., and subtypes, such as Transitive Verb, Intransitive Verb, etc., i.e., into a a hierarchy of types. It is possible to capitalize on this data classification computationally. In a preferred embodiment of the present invention, linguistic data may be specified as objects using well-known object oriented programming techniques. These techniques allow messages to be passed among objects regarding the manipulation of data pertinent to them and allow objects to inherit characteristics general to an overriding class of objects of which they are a member. The sublanguage dictionary organization allows particular features and processes to be indexed to entries in sublanguage dictionaries independent of any core dictionary entry for the same word, thus allowing the sublanguage use of that word to be domain-specific. Linguistic features and processes common to entries on several levels of the hierarchy can be used at all levels, thus streamlining maintenance of the linguistic analysis.

In its relational aspect, the Dictionary Database can allow access to and comparison of dictionaries on parallel levels in the hierarchy. Identical entries from the user dictionaries of several users in the same domain or group can be merged and "promoted" from user dictionaries to a higher-level dictionary upon application and satisfaction of certain reliability criteria. Utilities (described below) available to the Dictionary Maintenance Operator (DMO) allow investigation of relations between entries in user, subdomain, and domain dictionaries for the purpose of detecting and correcting overlapping or conflicting entries.

As illustrated in Fig. 4, the Dictionary Control Module 13

has two primary functions, Dictionary Selection and Dictionary Maintenance. In its Dictionary Selection capacity 13a, it determines the selection of core and sublanguage dictionaries to be used by the Machine Translation Module 20, based on the control data provided by the Recognition Module 12. The Dictionary Selection function includes an Inferencing Engine which assesses cover page or header data from the Recognition Module 12 and determines the dictionary selections to be supplied to the Machine Translation Module 20.

Dictionary Selection is exploited at the time of translation processing. It applies certain selection and ordering algorithms to the cover page or header data it receives from the Recognition Module 12 in order to determine the appropriate core and sublanguage dictionaries to be used for the translation of a given text. For example, the data can contain information as to core language pair, subject domain, correspondence type, and social indicators ("Mr." or "Mrs.", job titles, etc.), all of which may be used by the algorithms of the Inferencing Engine to select applicable sublanguage dictionaries. When each sublanguage dictionary is set up, specifications as to its usage parameters are indexed in the Dictionary Control Module 13.

Many different approaches to the sublanguage dictionary selection and ordering algorithms may be used depending upon the level and type of data obtainable from the cover page and other aspects of the overall system. For example, at a simplest level, the selection algorithms of the Dictionary Control Module 13 can designate sublanguage dictionaries which directly correspond to one or more variables of the cover page or header, e.g., the sender's name, the "communication type", the subject matter ("re"), etc.

A more developed sublanguage dictionary selection algorithm can make Boolean inferences about the context of the input text based upon the cover page data. For example, the user's group or relative social status can be determined by searching stored group lists or comparing the titles of the sender and recipient. This is particularly important for languages where correct terms of speech and address are dependent upon the context of a communication or the relative status of the parties. To illustrate, "IF the sender is a higher-level employee of a travel agency, AND the recipient is a lower-level employee of a hotel group, AND the communication is a "travel advisory", THEN USE Sublanguage Dictionary A23 for priority.1 lexical/grammar entries and Sublanguage Dictionary Z4 for priority.2 entries, OTHERWISE default to Core entries".

Dictionary Maintenance

The Dictionary Maintenance Module 13b enters, tracks, sorts, indexes, and maintains word entries in the multiplicity of core and sublanguage dictionaries. The Dictionary Maintenance Module 13b includes an interactive User Maintenance capability 13b-1 with an Input Editor for creating temporary "scratch" entries in user dictionaries, and a DMO Maintenance capability 13b-2 with a programmed Dictionary Maintenance Utility for updating dictionary entries based on data analyzed and supplied by more general DMO Assistance Utilities.

A user dictionary may be created or initialized at the time of subscription by a new user to the machine translation service. In the Interactive Mode, the new user may be prompted to provide samples

of typical texts expected to be submitted for translation, as well as individualized words for a thesaurus of terms relevant to that user. Later, the Input Editor of the User Maintenance Module 13b-1 may prompt a user to provide basic information that contribute to the maintenance of the user's dictionary. The Input Editor may be invoked in a variety of contexts. For example, during a translation session, the Input Editor may create scratch entries for the user's dictionary upon encountering unfamiliar words or phrases in the input text. During a dictionary-building session (outside of a translation session), the Input Editor may create scratch entries from a list supplied by the user. During a dictionary-maintenance session (outside of a translation session), the Input Editor may present the user with the contents of his or her personal dictionary for confirmation and updating. In all three contexts, the Input Editor draws upon algorithms designed to maximize the user's knowledge of the relationship of the words entered to one or more domains associated with the user, while requiring a minimum of user knowledge of linguistic principles or the structure of the Dictionary Database.

The user may also be offered a choice of how elaborate the interaction is to be. For example, the user may choose to spend the time necessary to answer numerous questions posed by the Input Editor about the syntactic and lexical properties of the new word. Alternatively, the user may choose an abbreviated option designed to provide only the linguistic information essential to a rudimentary translation of the word. The user may also be given the option of not creating a new entry for a particular word but settling for offering an acceptable substitute word or expression, or passing the source word into the target text untranslated. All these choices of interaction result in the creation of records of the interaction for

later examination by the DMO.

As illustrated in Fig. 5, those interactions that do call for creation of a scratch entry may do so by reference to similar entries or synonymous words already present in the Dictionary Database. Upon encountering an unfamiliar word or phrase, the Input Editor may ask the user whether the word is a domain-specific usage. If so, the user is prompted to name an appropriate domain or may be presented with a list of domains established in the Dictionary Database from which to choose. Then the Input Editor prompts the user for a one-word synonym of the unfamiliar word. The sublanguage dictionaries in the appropriate domain(s) are searched for an entry for that synonym. If one is present in a related sublanguage dictionary, it is imported to the user's dictionary. If not, the user may be prompted for additional synonyms and the process is repeated. If a synonym is not found in a domain-specific dictionary, the core dictionary may also be searched for the synonymous term.

If an appropriate synonym is found, a "copy-cat" entry is created for the new word in the user's dictionary, using the new word as the indexing name of the new entry and the content of the synonymous entry as its content. The user may be given the choice of using the content "as is", and the word is then translated in the manner specified by the content of the "copy-cat" entry. If the user does not want to use the entry "as is", the Input Editor may prompt the user for information regarding the new word's syntactic and lexical characteristics, to ascertain whether it is similar in those respects to the synonymous word from which the entry content was copied. The word may then be translated if the interaction results in an enabling specification of the new word. If not, the scratch

entry is maintained for later review by the DMO, and the user can choose to offer a substitute word or expression or pass the source word into the target text untranslated.

Figs. 7A - 7D illustrate an example of the creation of scratch word entries in user dictionaries and their promotion to a higher-level subdomain or domain dictionary. Fig. 7A is a schematic representation of the creation of an entry for the word "mouse" (i.e., a computer peripheral device) in the personal dictionary of User A, based on a similar entry in the personal dictionary of User B. Both users are members of the sub-domain "Computer". The personal dictionaries of both users are nested within the core dictionary of the language pair English-to-Japanese. Fig. 7A illustrates this process for a transfer-type machine translation system, however, it can be implemented for an interlingua system in an equivalent manner.

User B's personal dictionary contains the three types of entries necessary for transfer: an English monolingual entry (labeled "E_word"); an English-to-Japanese bilingual transfer entry (labeled "EJ_word"); and a Japanese monolingual entry (labeled "J_word ; 'computer mouse'"). These entries contain specifications uniquely labeled to refer to the rules and features that are pertinent to the word's grammatical functions and linguistic characteristics. Grammatical functions are specified by frame references, which refer to general rules for linguistic types (nouns, verbs, etc.). The inclusion of such references invokes the operation of the relevant rules by inheritance from files available for use by the MT system as a whole. Linguistic features are characteristics unique to individual words, and their values are supplied within the entries

themselves.

For instance, the English monolingual entry illustrated in Fig. 7A contains a NounType frame reference (E-NType-1) and three features (E-NFeatX, etc.). The bilingual entry contains a transfer-rule frame reference (EJ-NType-M) and the specification of the translation of the word "mouse" (the Japanese translation is represented here as **). The Japanese monolingual entry contains the frame reference J-NType-R and three features. The two monolingual entries contain a group specification, i.e., group = comp (Computer).

Upon encountering the new word "mouse" in text input to the system by User A, the Input Editor may interact with the user as follows:

Q: Is this a domain specific usage?
R: Yes.
Q: What is the domain?
R: Computers.

With this information, the Input Editor scans the personal dictionaries of other users in the "computers" sub-domain and finds an entry for "mouse" in the dictionary of User B. The Input Editor performs certain checks to ascertain whether this entry is in fact pertinent to the domain in question, including a search for group type, e.g., whether User B's entry for "mouse" contains the specification group = comp. The group specification for the word "mouse" may be necessary in addition to User B's membership in the sub-domain "computer", since User B may be a member of other groups and domains as well. Additional checks may be performed to determine

the relevance of the entry, in the form of further questions posed to User A. In this illustration, for example, the Input Editor could ask User A whether the word "mouse" is a noun (inferring Noun status from the names of the frame references and features within the entries), whether it had a positive value for NFeatX, etc.

Upon determination that the entry for "mouse" found in User B's dictionary is a suitable match for User A's purposes, entries for User A's dictionary are created by making a copy of all relevant entries from User B's dictionary, and a record of the transaction is made for later perusal by the DMO. The copying process allows the generic functions of the noun to be inherited by User A's entry through the frame references (illustrated by the small box in the top center of Fig. 7A), as well as specifying the featural characteristics unique to the word "mouse".

If no entries for the word are found in the dictionaries of users within the "computer" domain, the Input Editor may scan for entries in both lower- and higher-level dictionaries, including the core dictionary. Fig. 7B illustrates this broader scanning process. As the search moves farther from the domain specified by User A, the Input Editor can proceed with greater caution in selecting candidate source entries for building User A's scratch entries by further requiring checks and caveats to the user, and can also include a notation of the heightened caution in the DMO record. Fig. 7C illustrates an example where the same entry is found in another user dictionary in another domain.

If no entries for the same word are found in the available dictionaries, the Input Editor may ask for a synonym for the word.

In this example, User A may respond with "pointer" as a synonym for "mouse". The system can then scan the various levels of dictionaries for entries for the synonym. If an entry for the synonym is found, e.g., in User F's personal dictionary, the Input Editor can pose a series of questions to User A based upon inferences made from the contents of the found entry or entries, as outlined above. If a determination is made that User F's synonymous entry is a suitable match for User A's purposes, a "copy-cat" entry is created, with the entry label of User A's term (e.g., "mouse") and the contents of the User F's entry relevant to "pointer". Once again, a record of this transaction is made for further oversight by the DMO. Fig. 7D illustrates the "copy-cat" entry creation process.

The addition of scratch entries through the Input Editor of the User Maintenance Module 13b-1 into user dictionaries provides the MT system with a pre-screened corpora of words as to which basic information on linguistic features and domain (sublanguage) relations have been supplied. Thus, the Dictionary Database can cumulate and evolve over time along with actual words and usages encountered in texts or supplied by users.

An important feature of the present invention is the capability to move word entries from a lower-level (user, subdomain/group, or domain) dictionary into a higher-level (subdomain/group, domain, or core) dictionary, when those entries have met certain tests of linguistic completeness and more general usage that indicates the desirability of inclusion in a higher-level dictionary. A language is in a constant state of evolution as new words and usages are adopted by individuals and groups and then gain currency through larger groups and the society as a whole. The

movement of entries into more general dictionaries requires review and monitoring by a dictionary maintenance operator (DMO) trained in linguistics and/or translation of new words and expressions, in order to ensure that inaccurate translations or corruptions of the Dictionary Database do not occur. The present invention provides for certain automated utilities in the DMO Maintenace Module 13b-2 which assist the DMO in the movement of word entries to higher-level dictionaries. If word entries can satisfy certain tests indicating that their upward (more general) movement can be done with a high degree of reliability, the DMO assistance utilities may perform the movement of such word entries on a fully automated basis. Such programmed utilities for DMO assistance or fully automated maintenance open the way for substantially computerized management of very large dictionary databases, which will enhance the accuracy, performance, and utility of machine translation systems.

The manifold tasks of Dictionary Maintenance are fundamental to the operation of the machine translation system described herein. The tasks of the DMO are fulfilled, ultimately, by adding new dictionary entries and deleting or modifying already existing ones. The choice of which entries are to be added or deleted, which aspects of entries are to be modified, and which dictionaries and sublanguage dictionaries are to be affected by these changes is made by the DMO based upon data on usage and effectiveness of entries as derived by employing the programmed assistance utilities.

As illustrative examples, the DMO Maintenance Module 13b-2 may include basic utilities to aid in extracting and organizing word entries from the underlying corpora, such as: conversion of compiled

dictionary entries (including grammar rules) to text files for display, for ease of perusal and manual editing; automated creation of the records mentioned above and presenting them as text files; detection of errors in the structure of dictionary entries and presenting faulty entries as text files; extracting and displaying a list of all lexical entries by entry name only (a "shelf list" of entries in the dictionary); extracting and displaying a list of all bilingual lexical entries by entry name and translation only (including syntactic category for disambiguation), i.e., a bilingual "shelf list"; and/or detection of missing links in inheritance hierarchies, which would prevent access to a higher-level dictionary without such link(s).

This set of extraction and display utilities can assist the DMO in maintaining a level of consistency and exhaustiveness of control in Dictionary Database monitoring. A primary focus of the present invention is identifying, creating, maintaining, and using sublanguage dictionaries suited to the lexical and linguistic idiosyncracies of groups of users. These aspects require the ongoing assessment and maintenance of the relations between and among sublanguage dictionaries, based upon fluid word usage patterns of members of the same domains or groups.

Maximal efficiency in performing these dictionary-relation tasks call for sophisticated utilities to be available to the DMO. Such utilities may include the capability to keep track of all instances of entries of a new word by users in the same domain. A frequency-of-use utility can determine the frequency of use, number of entries, and identity of a preferred synonymous entry. A further utility can present such data to the DMO for examination for possible

"promotion" from user dictionaries to a domain dictionary. A sublanguage-selection utility can perform an analysis of the above-mentioned records and display patterns of use (e.g., frequency and consistency of usage), indexed to individual users and groups of users, to assess the accuracy of the sublanguage dictionary selection process. A homograph entry utility can identify and display entries of homographs of new entries in user dictionaries in a given domain, to display to the DMO for analysis in determining the optimal formulation of the homographic entries, and possible collapsing of entries into higher level dictionaries. A quality assurance utility can display newly created scratch entries for quality assurance checks or for promotion to a higher-level dictionary of approved entries.

As aids in maintaining the consistency of the linguistic and semantic feature network of the system, utilities may also be provided for: finding and displaying entries containing a certain feature-value pair, e.g., displaying entries containing the feature [+liquid]; displaying a shelf list of entries containing a certain feature-value pair; and displaying the organization of features to assist in tracking feature-assignment errors.

The DMO Assistance Utilities may also employ algorithms to perform cross-dictionary comparisons, concordances, integration, differentiation, statistical matching, cluster analysis, etc., in order to resolve matching, conflicting or overlapping entries in different dictionaries. For example, the utilities may be used to scan the dictionaries of users from the same domain or group to see whether any word entries may be collapsed into a more general word entry and "promoted" from the user dictionaries to the domain

dictionary. An example of such entry promotion is shown in Fig. 6.

Criteria for a sufficient level of similarity among entries in subordinate dictionaries can be measured using the statistical or numerical algorithms indicated above. Such measures and others may also be employed to determine which characteristics of entries are general and thus suited for inclusion in the entries to be created for the higher level dictionaries, and which characteristics are idiosyncratic to the users whose dictionaries are the source of the entries. Thus, promotion of entries to higher level dictionaries need not involve their erasure from the user level. The higher level entries are promoted with only the applicable general characteristics. Idiosyncratic characteristics, if any, may be kept in the user dictionaries. Entry promotion can occur between adjacent levels within the dictionary hierarchy.

The DMO Assistance Utilities may also measure co-occurrences of words and terms in the input texts of individual users or groups of users to determine group membership and relations between groups and members, and to infer their characteristics based upon characteristics recorded for similar groups or members, or to derive general sublanguage patterns for creation of master (superordinate) sublanguage dictionaries. Such analyses may also yield lists of key words for a thesaurus that is used to select a sublanguage dictionary appropriate to certain users.

With information contributed by the DMO Assistance Utilities, the DMO can create lists of words and phrases to enter, delete, or move from one sublanguage to another. Similarly, the DMO can employ data supplied by feature extraction utilities to arrive at lists of

features to alter in the appropriate lists of entries. The DMO may create and alter entries by inputting the data to a text file and compiling the file into the Dictionary Database. However, a more sophisticated and efficient set of utilities can be provided which automates the creation and modification of lexical entries, including source entries, target entries, and source-to-target entries in a dictionary.

In summary, the Dictionary Control Module allows the overall machine translation system to possess a very fluid and highly granular sublanguage capability. The sublanguage capability is developed and cumulated over time based upon the encountered words and identified preferences of actual users, user groups, domains, or fields. The multiple sublanguage dictionaries are like the listings of synonyms and alternate phrase usages in a real-world dictionary, except that the entries can change along with usage, and the capacity for domain-specific usages is virtually limitless. Computational power in the manipulation of the multiplicity of sublanguage dictionaries replaces the need to rigorously define an overall set of sublanguage patterns for a given domain. Horizontal expansion of sublanguage capability thus replaces vertical definition.

Output Formatting and Transmission

As shown in Fig. 1C, the Output Module includes a Page Formatting capability 31 driven by formatting instructions extracted from the cover page or header for the translated text by the Recognition Module 12, and a Sending Interface 32 which transmits the formatted output text via telecommunications link B to recipients at

their addresses as extracted from the cover page and supplied by the Recognition Module 12.

Once the input text has been translated into target language text, the Page Formatting capability of the Output Module 30 composes the translated text into a desired page format based upon the formatting information designated on the cover page. For example, for English-to-Japanese translation, the output Japanese text may be formatted as "left-to-right" horizontal lines of kana, or as "right-to-left" vertical lines of ideographic characters. The page format may also be designated for "page-by-page" translation, wherein the formatting program takes into account the compression ratio between the source and target text. For example, as illustrated in Fig. 3, English text is typically more spatially expansive than ideographic text, so that an 8.5" x 11" input page of English text may be reformatted on the same size page with Chinese characters of suitably larger point size and interline spacings. Correspondingly, a typical 15.2 cm x 25.6 cm page of ideographic text may be reformatted as an 8.5" x 11" page of English text, or an 8.5" x 11" or A4-size page may be reformatted as an 8.5" x 14" page.

The formatting program may also implement a footnoting function, as shown in the section "F" in Fig. 3, providing footnotes for ambiguous phrases of the input text by replicating their original source language text (indicated, for example, by a single asterisk) and/or providing alternate translations in the target language (indicated by double asterisks). The source language phrase and/or alternate translation are provided by the Machine Translation Module 20 by flagging an ambiguous word or phrase which could not be resolved in the translation processing. Other well-known page

formatting functions, e.g., margins, page layout, columns, replication of non-translatable graphic images, etc., may also be performed by the Output Module 30.

When the formatted output document is ready for output transmission, the Sending Interface 32 becomes operative to generate the command signals for controlling the corresponding output devices and sending the output document as electronic data signals to the respective devices through the telecommunications link B. The output devices can include a telephone fax/modem board, a printer which may also be coupled with a page facsimile transmission machine and/or an automatic mailing machine (for mailing hard copy), or a network interface for sending the output data to a recipient's electronic address on a network. As shown in Fig. 2, the cover page may designate a plurality of recipients in a plurality of target languages and located at a plurality of addresses. The Sending Interface 32 generates and routes the appropriate forms of output data to each recipient. For example, if each recipient is designated to receive a fax transmission and a printed copy, the Sending Interface routes the data through the fax/modem board to each recipient's fax number and also activates the printer and collation of pages for automatic mailing.

General Telecommunications Use

The above-described machine translation system can be adapted to and installed as a resident utility or service in telecommunications systems or networks, such as private and public networks and gateway companies, telecommunications companies, and bi-

or multi-lingual information service providers. The input to the telecommunications system is preferably in the form of electronic text transmissions in the near term. However, with further development, input in the form of graphics (facsimile) data and even speech can be captured, scanned, and converted to intermediary text for translation processing and output formatting and transmission in any form. As the Dictionary Database develops in depth of coverage of particular sublanguages and in breadth of coverage of many different domains, the system will acquire the accuracy and capability to handle communications over a wide range of fields satisfactorily. Mass storage and inexpensive processing power and speed can be effectively utilized to handle many different language pairs, technical fields, domains, user groups, and individual users for near-simultaneous translations in a host of languages.

For near term use, the machine translation system is particularly suitable for translating electronic text for E-mail, electronic bulletin boards, and information services in telecommunications networks. As described above, an Interactive Mode may be provided through a telecommunications program or a network system to interact with an online user inputting text to be translated and transmitted on the network. In this mode the user may be prompted to fill in cover page fields or create and maintain User ID files, to aid in dictionary selection, or update the Dictionary Database, to enable the system to translate new words encountered in the input.

For electronic text input and output in different languages, it is desireable to have a standardized interface to the many different character code conventions used throughout the world.

A universal character code convention has been developed by the Unicode Consortium, Mountain View, California. The Consortium includes IBM, DEC, Apple, and other major American computer companies. The Unicode set is a 16-bit character code set that is mapped to the major character code conventions of the world, including the major Roman alphabetic systems and Asian character systems. For example, the Unicode set is mapped to the Han character sets of the major industry and national standards used in China, Japan, Korea, and Taiwan. Thus, a Unicode character converter module can be employed as the standardized interface for electronic text in a telecommunications system.

Text input can also be scanned in from printed pages or from transmissions via a fax/modem. The system's Recognition Module is used to convert such scanned page image data into machine-readable text. Currently, off-the-shelf programs are available for English alphanumerics and Japanese kana. Future developments in character recognition programs for other character sets, such as Chinese Han and Japanese kanji characters, and even handwritten characters, can be expected to further the development of page-oriented translation systems.

CLAIMS:

1. A machine translation system for translation of input texts sent from a plurality of different users, wherein each of said users may have a preferred sublanguage of text terminology used in the user's input text out of a plurality of possible sublanguages, and wherein a perferred sublanguage of a user is determinable from one or more parameters of the user's identity, said machine translation system comprising:

(a) a receiving interface for receiving a series of translation jobs to be translated in sequence, each translation job comprising an input text and accompanying control input including user ID data identifying a user sending the input text, wherein said receiving interface includes means for identifying the input texts and accompanying control inputs for each one of the series of translation jobs and for queueing the input texts for translation in sequence;

(b) a machine translation module for performing machine translation of the input texts in sequence by translating each input text in a source language to an output text in a target language using a dictionary database containing entries for words of the source and target languages;

(c) a dictionary database including a core dictionary containing entries for generic words of the source and target languages, and a plurality of sublanguage dictionaries each containing entries for specialized words of a respective one of a plurality of sublanguages handled by said machine translation system for the source/target languages;

(d) a recognition module including a memory section for storing a plurality of user ID files each of which contains previously stored user sublanguage preference information which is indexed to

user ID data for each respective user of said plurality of users, said user sublanguage preference information being indicative of a sublanguage of text terminology preferred by the respective user for translation of an input text from that user, wherein said recognition module is responsive to the user ID data received by said receiving interface to retrieve the user sublanguage preference information stored in the user ID file indexed to the user ID data;

(e) a dictionary control module responsive to the particular user sublanguage preference information retrieved by said recognition module for selecting a corresponding one of the plurality of sublanguage dictionaries of the dictionary database, and for causing the machine translation module to use the selected sublanguage dictionary along with the core dictionary for translation of the particular input text of each respective user; and

(f) an output module for outputting text in the target language translated by the machine translation module for each one of the input texts, whereby said receiving interface identifies for each translation job in sequence the input text to be translated and the control input including user ID data identifying a particular user sending the particular input text and forwards the particular user ID data to the recognition module, the recognition module retrieves the particular user sublanguage preference information from the user ID files indexed to the particular user ID data and forwards it to the dictionary control module, and the dictionary control module selects the preferred sublanguage dictionary out of the plurality of sublanguage dictionaries that corresponds to the particular user sublanguage preference information for use by the machine translation module.

2. A machine translation system according to Claim 1,

wherein said dictionary database contains a plurality of core language dictionaries corresponding respectively to a plurality of source/target languages for machine translation by said machine translation module, wherein said control input for each translation job includes a source/target languages control input indicative of a selected source/target core language applicable to the accompanying input text, and said dictionary control module is responsive to the source/target languages control input identified by said receiving interface and causes said machine translation module to use a corresponding source/target core language dictionary in performing translation of the input text.

3. A machine translation system according to Claim 1, wherein said dictionary control module contains an inferencing program for selecting an applicable sublanguage dictionary based upon said sublanguage preference information indicating one or more parameters of a user's identity including title, sex, company, job position, address, user group, and subject matter.

4. A machine translation and telecommunications system comprising:

(a) a receiving interface for receiving via a first telecommunications link an input text in a source language accompanied by a control input including a first predefined field containing an address of a receipient to receive output text translated to a target language and a second predefined field containing a source/target language control input indicative of a selected source/target language pair for translation applicable to the input text from among a plurality of possible source/target language pairs;

(b) a machine translation module capable of performing

machine translation of an input text in a source language to an output text in a target language using a dictionary database containing entries for words of the target language corresponding to words of the source language;

(c) a dictionary database containing a plurality of source/target language dictionaries, each containing entries for generic words for translation between a source and target language pair;

(d) a dictionary control module responsive to the source/target language control input for selecting a source/target language dictionary of the dictionary database which is applicable to the input text, and for causing the machine translation module to use the selected source/target language dictionary in performing translation of the input text to the designated target language; and

(e) an output module responsive to the address of the control input for outputting translated text in the target language generated by the machine translation module and automatically routing it to be sent to the recipient's address.

5. A machine translation and telecommunications system according to Claim 4, wherein the control input includes a third predefined field containing a sublanguage control input for selecting a sublanguage of a source/target language to be used for translation of the input text, said dictionary database containing a plurality of sublanguage dictionaries, each containing entries for specialized words of a sublanguage domain for translation within a source and target language pair, and said dictionary control module being responsive to the sublanguage control input for selecting a sublanguage dictionary of the dictionary database which is applicable to the input text, and for causing the machine translation module to

use the selected sublanguage dictionary in performing translation of the input text.

6. A machine translation and telecommunications system according to Claim 4, further comprising a recognition module coupled to said receiving interface for electronically scanning the control input and recognizing the address of the recipient designated in the first predefined field of the control portion, and said output module including a sending interface for sending the translated output text generated by said machine translation module to the address of the recipient recognized by said recognition module.

7. A machine translation and telecommunications system according to Claim 4, wherein said receiving interface includes a programmed interaction module for interactive input from a user through a user interface to said system via said first telecommunications link.

8. A machine translation and telecommunications system according to Claim 4, wherein said input text and translated output text are transmitted as electronic text, and said output module transmits the translated output text via a second telecommunications link.

9. A machine translation and telecommunications system according to Claim 8, wherein said system is installed as a resident utility or server in a telecommunications system or network.

10. A machine translation and telecommunications system according to Claim 8, wherein said receiving interface and said output

module include an electronic character code interface for receiving and sending electronic text in any of a plurality of electronic character coding conventions used by senders and recipients in different languages.

11. A machine translation and telecommunications system according to Claim 4, wherein said input text is received as graphics data, and said receiving interface includes a character recognition module for converting text content contained in said graphics data to machine-readable electronic text.

12. A machine translation and telecommunications system according to Claim 4, wherein said output module includes a graphics output module for converting machine-readable output text to graphics data for facsimile or other graphics transmission.

13. A machine translation and telecommunications system according to Claim 4, wherein said output module is configured to send the translated output text together with the input text to the recipient's address to allow verifying of the translation.

14. A machine translation and telecommunications system comprising:

(a) a receiving interface for receiving via a first telecommunications link an input text in a source language accompanied by a control input including a first predefined field designating an address of a receipient to receive output text translated to a target language and a second predefined field containing a source/target language control input designating a selected source/target language pair for translation applicable to the input text from among a

plurality of source/target language pairs;

(b) a machine translation module capable of performing machine translation of an input text in a source language to an output text in a target language, said module having a plurality of source/target language pair submodules any one of which can be operated for automatically performing machine translation in a selected source/target language pair;

(c) a control module responsive to the source/target language control input of the second predefined field for selecting a source/target language pair submodule for operating said machine translation module to perform machine translation of the accompanying input text in the source/target language pair to the designated target language; and

(d) an output module responsive to the recipient's address of the first predefined field for outputting the translated text in the target language generated by the machine translation module and automatically routing it for sending to the recipient's address.

15. A machine translation and telecommunications system according to Claim 14, wherein said source/target language pair submodules are provided with respective dictionary databases for performing translation in their respective source/target language pairs, and each dictionary database includes a plurality of sublanguage dictionaries containing entries for specialized words of respective sublanguage domains within the respective source/target language pair, wherein the control input includes a third predefined field containing a sublanguage control input for selecting a sublanguage domain of a source/target language pair that is preferred for translation of the accompanying input text, and wherein said control module is responsive to the sublanguage control input of the

third predefined field for selecting the preferred sublanguage dictionary of the dictionary database and operating said machine translation module to perform machine translation of the input text using the preferred sublanguage dictionary.

16. A machine translation and telecommunications system according to Claim 14, wherein said input text and translated output text are transmitted as electronic text, and said output module transmits the translated output text via a second telecommunications link.

17. A machine translation and telecommunications system according to Claim 16, wherein said system is installed as a resident utility or server in a telecommunications system or network.

18. A method of automatically translating text from a source language into a target language and sending the translated output text to a designated recipient via a telecommunications network comprising the steps of:

(a) sending input text as electronic text via a first telecommunications link to a resident utility or server on the telecommunications network, said electronic input text being in a source language and being accompanied by a control input including a first predefined field designating an address of a recipient addressable through the telecommunications network to receive output text translated to a target language and a second predefined field containing a source/target language control input designating a selected source/target language pair for translation applicable to the input text from among a plurality of source/target language pairs;

(b) accessing at the resident utility or server a machine

translation module capable of performing machine translation of an input text in a source language to an output text in a target language, said machine translation module having a plurality of source/target language pair submodules any one of which can be operated for automatically performing machine translation in a selected source/target language pair, and said machine translation being responsive to the source/target language control input of the second predefined field for selecting a source/target language pair submodule for performing machine translation of the accompanying input text in the source/target language pair to the designated target language; and

(c) automatically sending the translated output text as electronic text via a second telecommunications link through the telecommunications network to the recipient's address as designated in the first predefined field of the control input accompanying the input text.

19. A method of automatically translating and sending text via a telecommunications network according to Claim 18, wherein the electronic text is transmitted as E-mail, transmitted to or from electronic bulletin boards, or transmitted to or from information service providers via the telecommunications network.

20. A method of automatically translating and sending text via a telecommunications network according to Claim 18, further comprising the step of converting electronic input text sent to and from the resident utility or server in any of a plurality of electronic character coding conventions used by senders and recipients in different languages.

Abstract of the Disclosure

A machine translation and telecommunications system includes a machine translation [engine] module for translation of input text [from a source language to a] between any of a plurality of source and target languages, using a dictionary database including a plurality of core [dictionary] dictionaries and a plurality of sublanguage (domain) dictionaries usable for translation from a source to a target language, a receiving interface for receiving text input from any of a plurality of users, each text input being accompanied by control [information including user ID data indicative of] input designating an address of a recipient of the translated output and a selected source/target language pair and/or one or more sublanguages preferred [by a particular user] for the translation, and an output interface for automatically routing the translated text to the recipient's address. [, and a dictionary control module coupled to the receiving interface responsive to the user ID data indicative of a sublanguage preference of a particular user for selecting a corresponding sublanguage dictionary of the dictionary database to be used by the machine translation engine along with the core dictionary for performing translation of the particular user's text input. User dictionaries can be maintained and selected to enhance translation accuracy in the same manner. The dictionary database encompassing core, sublanguage (domain), and user dictionaries is cumulated for greater capability over time through the use of dictionary maintenance utilities for updating the dictionaries.] The system is particularly adaptable for sending and receiving electronic text to and from a resident utility or server which performs the translation and routing functions automatically on a telecommunications network.

FIG. 1

MACHINE TRANSLATION SYSTEM

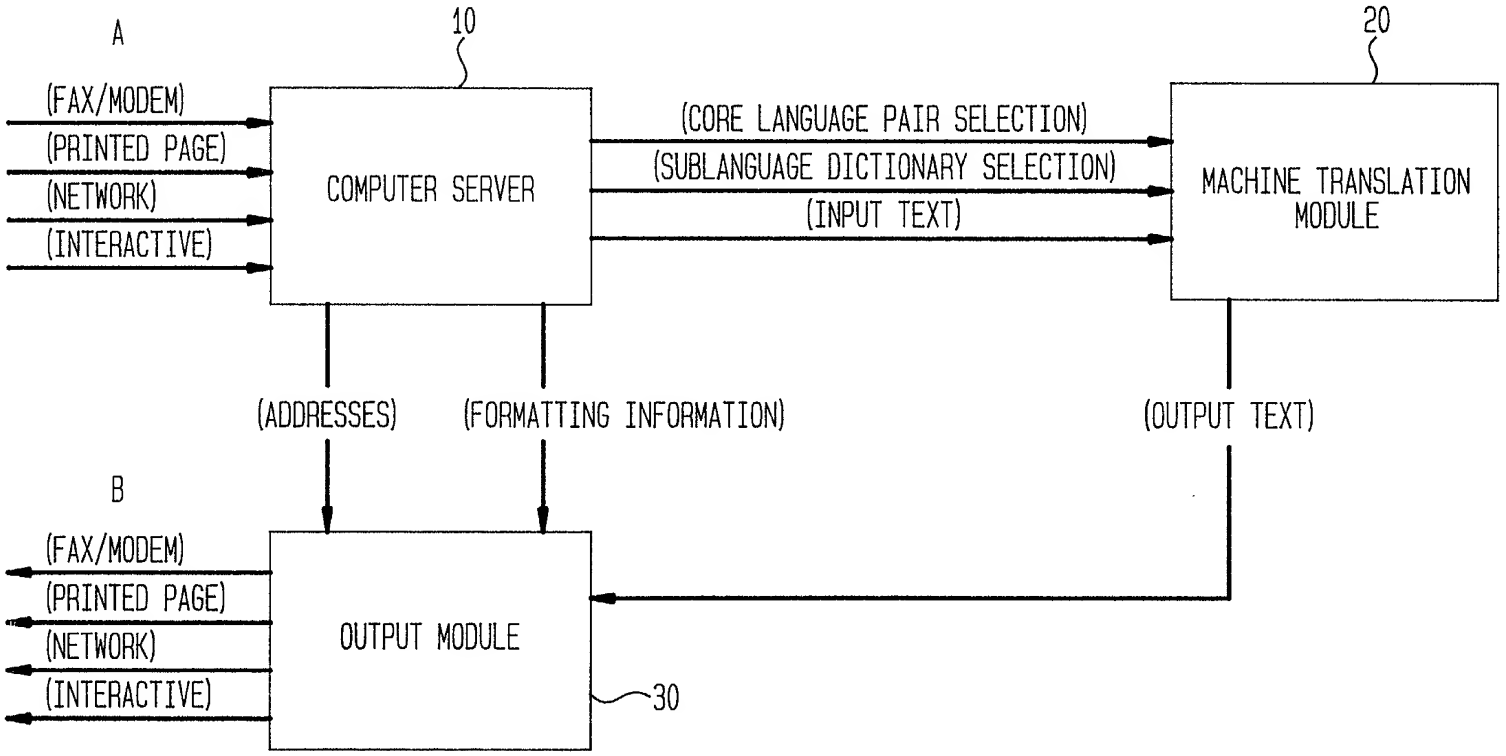


FIG. 1A

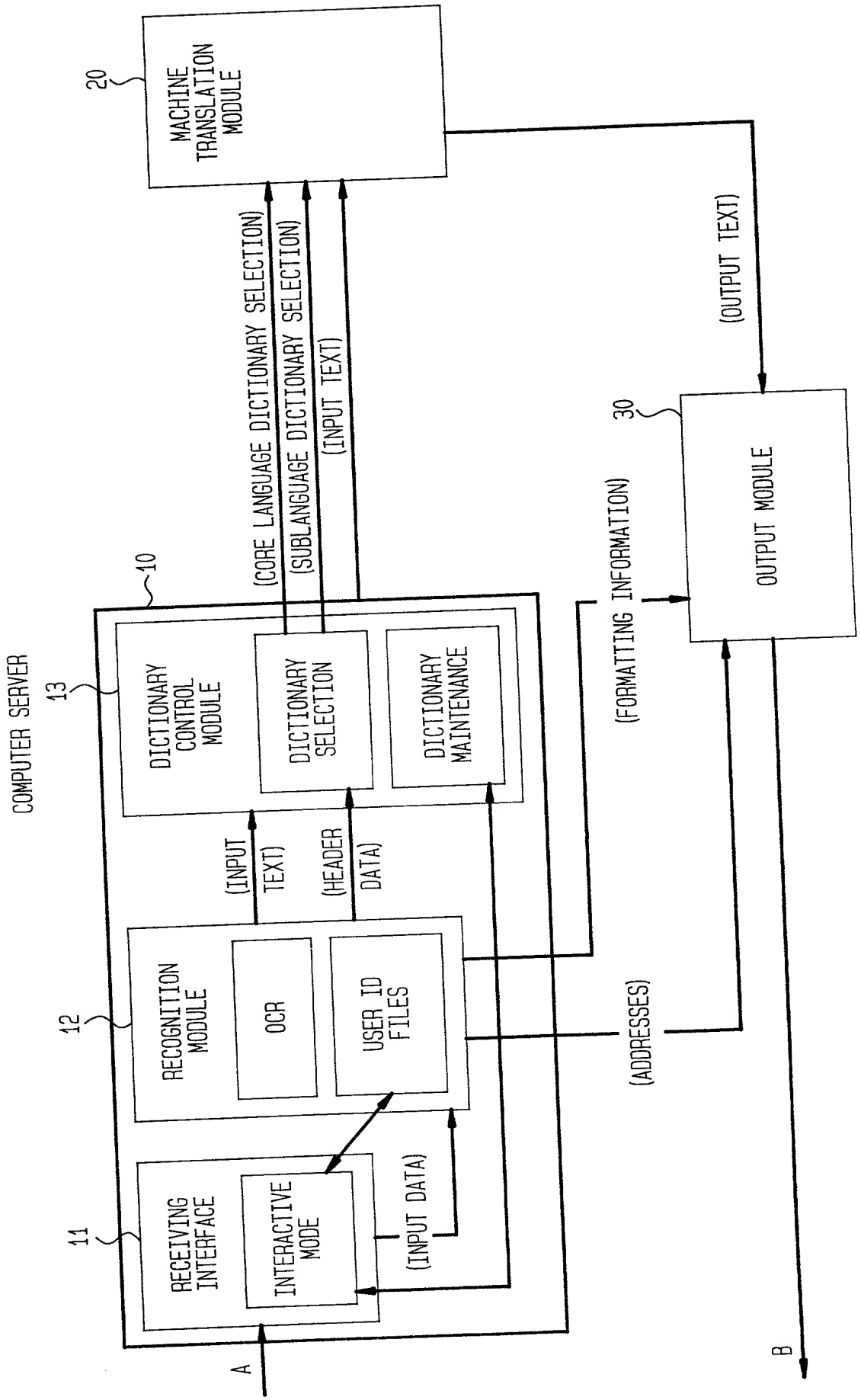
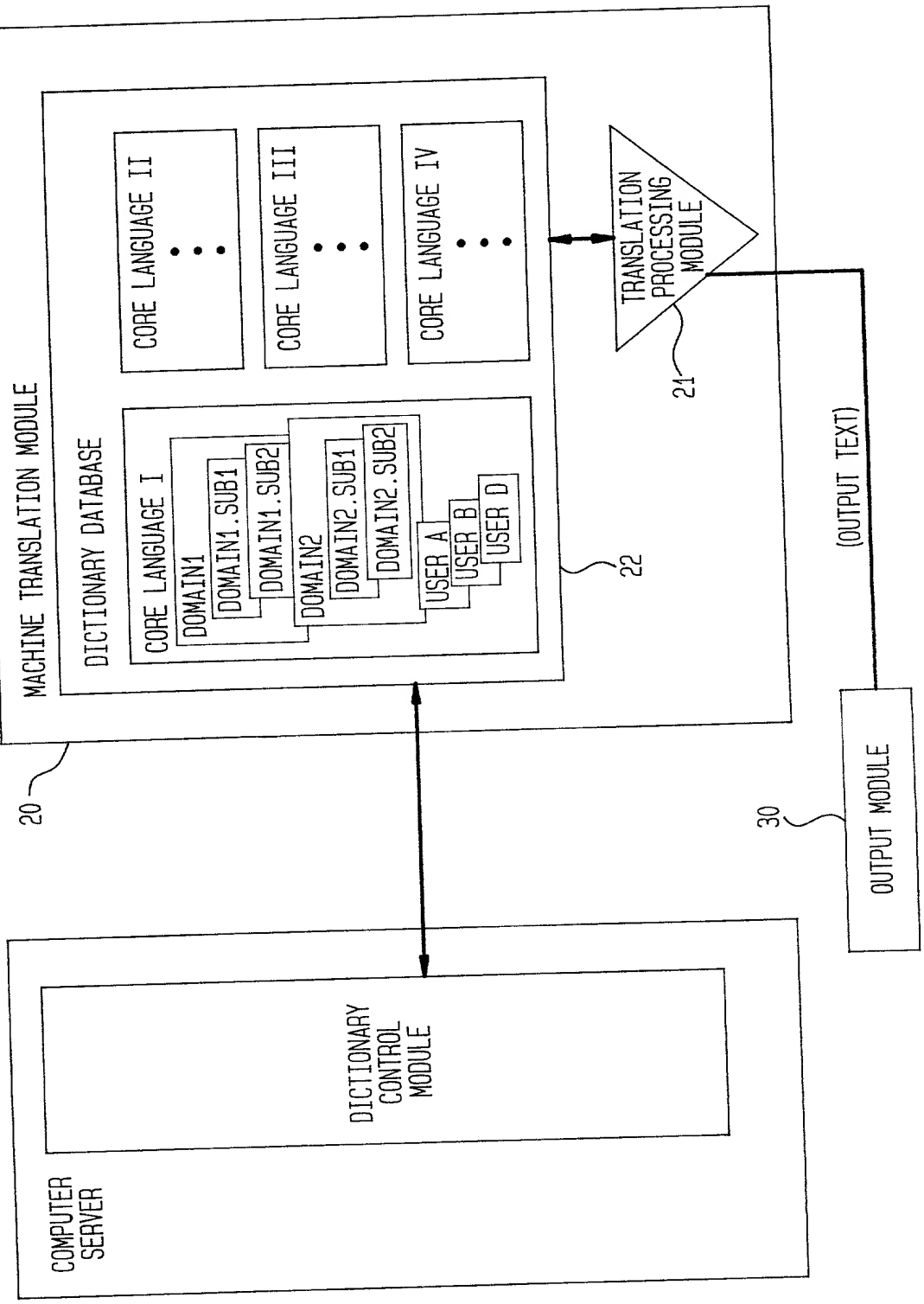

20
MACHINE TRANSLATION MODULE
(CORE LANGUAGE DICTIONARY SELECTION)
(SUBLANGUAGE DICTIONARY SELECTION)
(INPUT TEXT)
(OUTPUT TEXT)
30
OUTPUT MODULE
(FORMATTING INFORMATION)
10
COMPUTER SERVER
13
DICTIONARY CONTROL MODULE
DICTIONARY SELECTION
DICTIONARY MAINTENANCE
(INPUT TEXT)
(HEADER DATA)
12
RECOGNITION MODULE
OCR
USER ID FILES
(ADDRESSES)
11
RECEIVING INTERFACE
INTERACTIVE MODE
(INPUT DATA)
A
B

FIG. 1B

MACHINE TRANSLATION MODULE



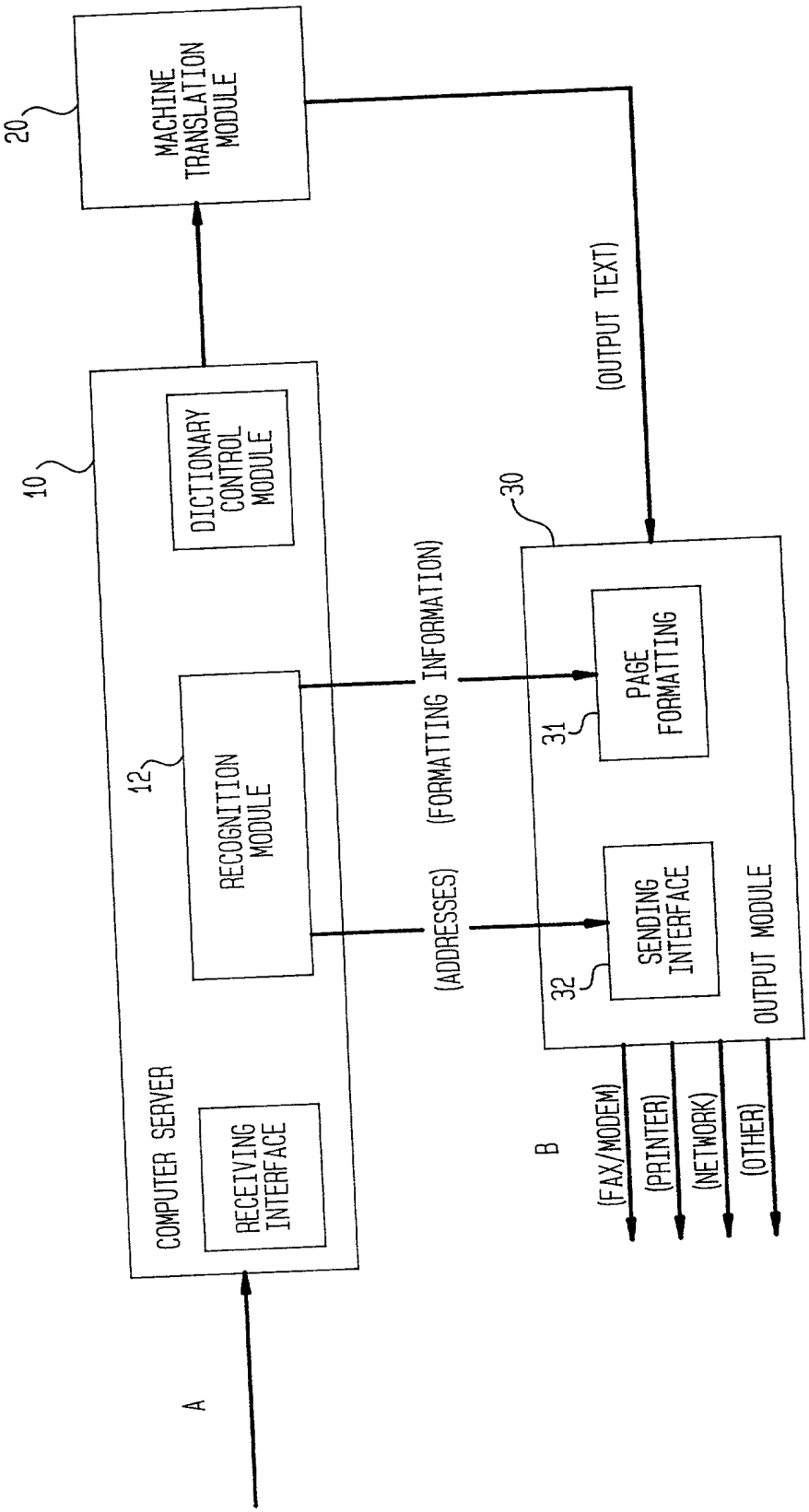


FIG. 1C

OUTPUT MODULE

FIG. 2

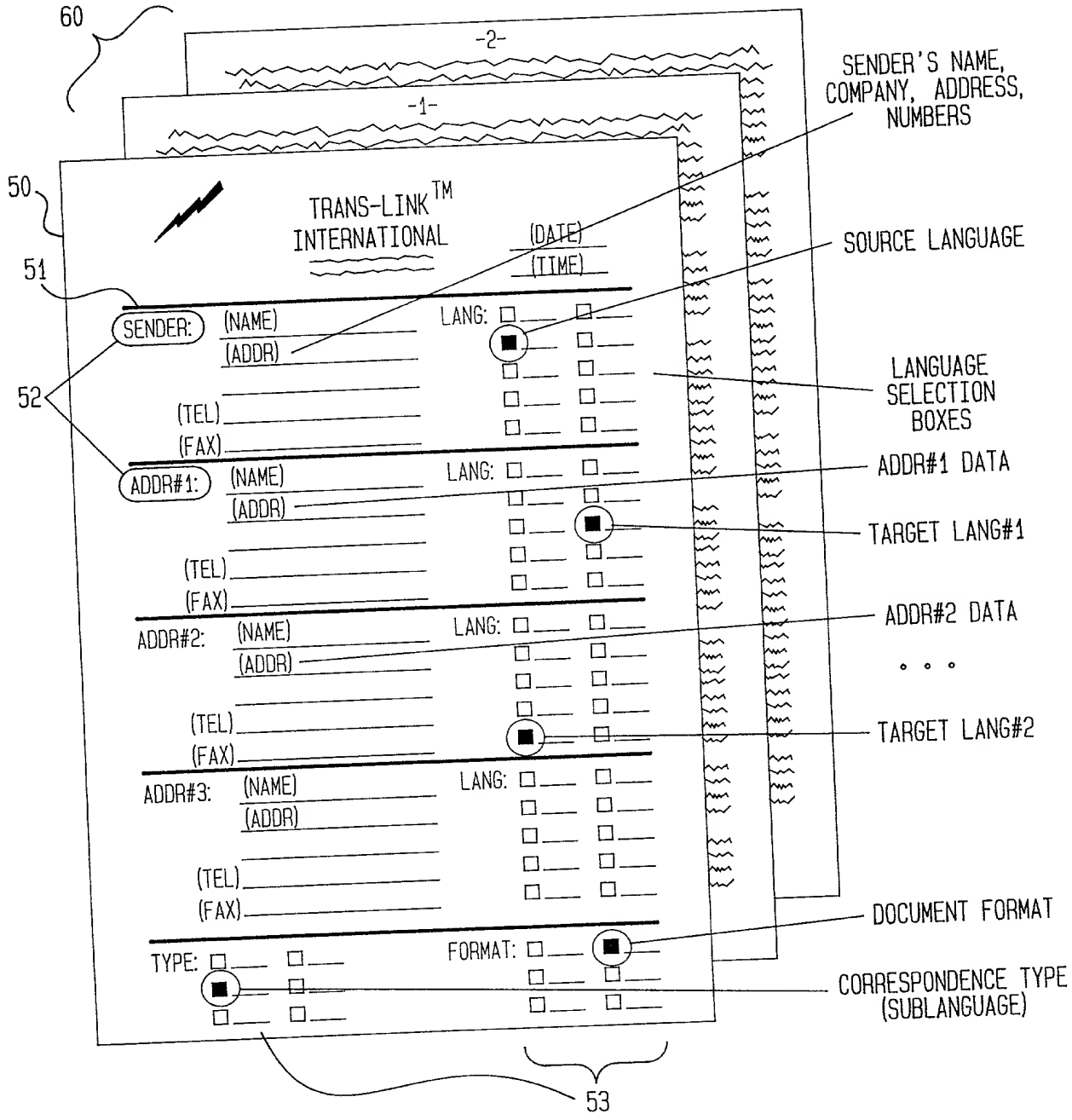
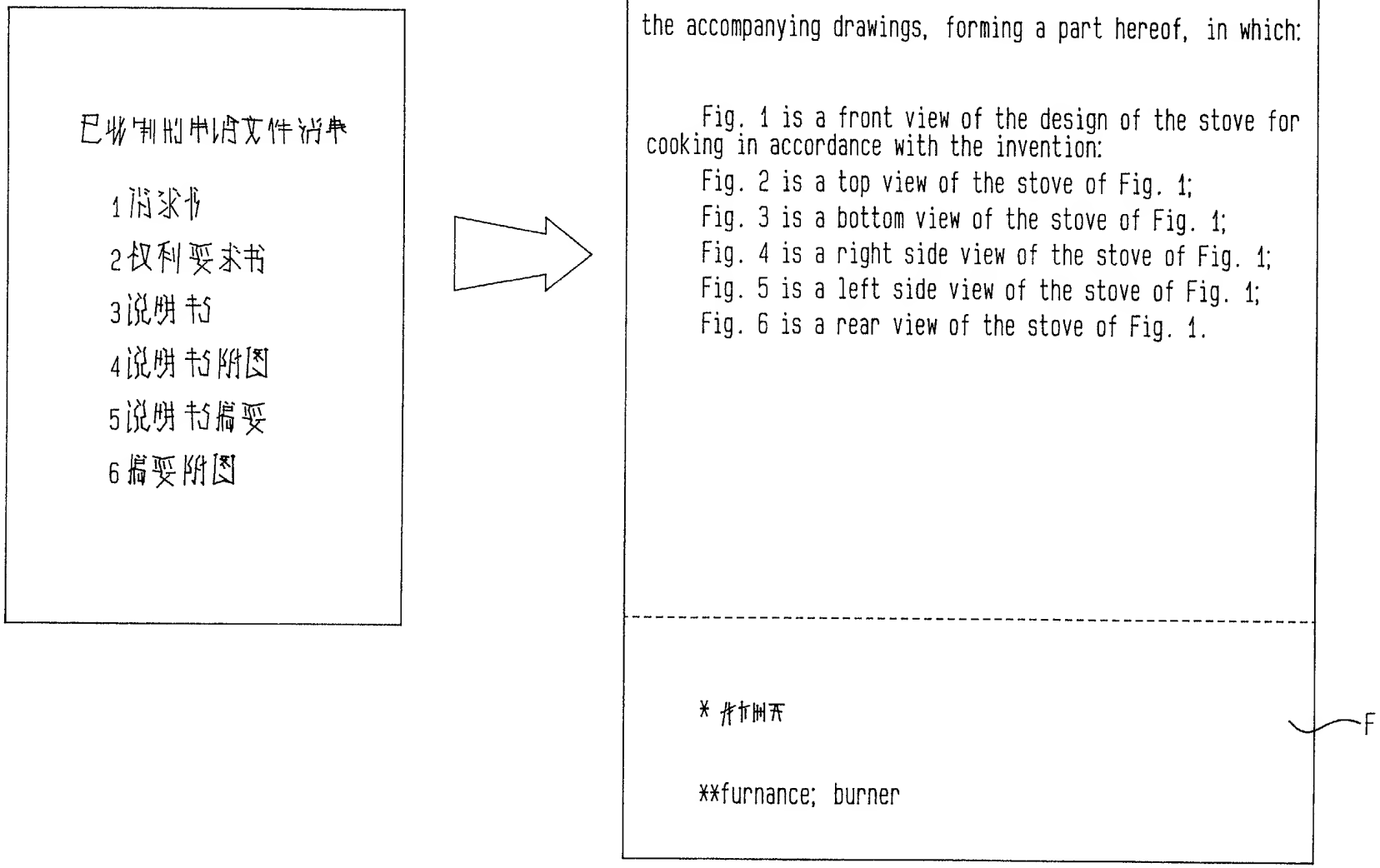

60
-2-
-1-
50
TRANS-LINK™
INTERNATIONAL
(DATE)
(TIME)
51
SENDER:
(NAME)
(ADDR)
(TEL)
(FAX)
LANG:
52
ADDR#1:
ADDR#2:
ADDR#3:
TYPE:
FORMAT:
SENDER'S NAME, COMPANY, ADDRESS, NUMBERS
SOURCE LANGUAGE
LANGUAGE SELECTION BOXES
ADDR#1 DATA
TARGET LANG#1
ADDR#2 DATA
. . .
TARGET LANG#2
DOCUMENT FORMAT
CORRESPONDENCE TYPE (SUBLANGUAGE)
53

FIG. 3


已收到的申请文件清单
1 请求书
2 权利要求书
3 说明书
4 说明书附图
5 说明书摘要
6 摘要附图
the accompanying drawings, forming a part hereof, in which:
Fig. 1 is a front view of the design of the stove for cooking in accordance with the invention;
Fig. 2 is a top view of the stove of Fig. 1;
Fig. 3 is a bottom view of the stove of Fig. 1;
Fig. 4 is a right side view of the stove of Fig. 1;
Fig. 5 is a left side view of the stove of Fig. 1;
Fig. 6 is a rear view of the stove of Fig. 1.
* 作为例子
**furnance; burner
F

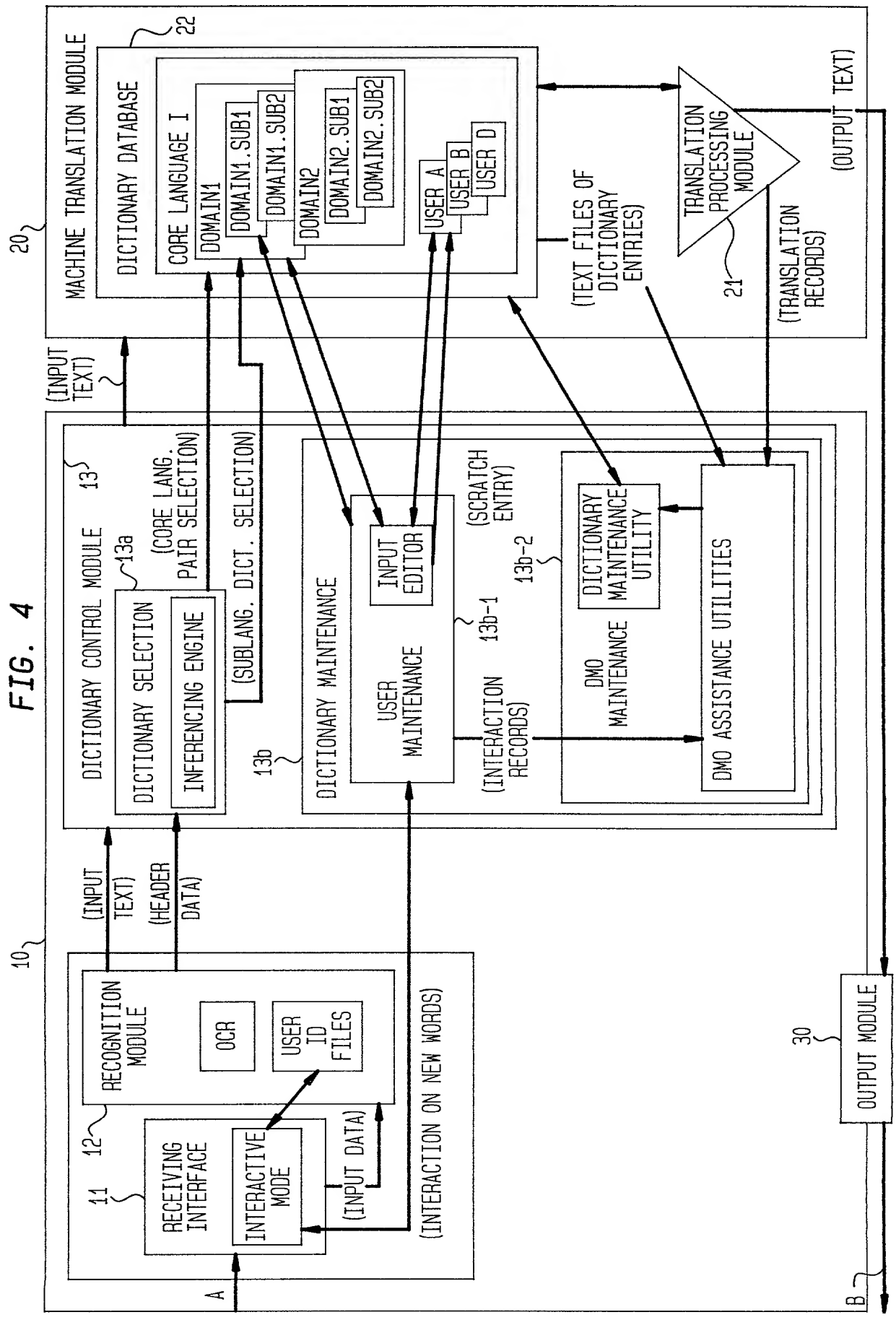

10
20
MACHINE TRANSLATION MODULE
DICTIONARY DATABASE
22
CORE LANGUAGE I
DOMAIN1
DOMAIN1.SUB1
DOMAIN1.SUB2
DOMAIN2
DOMAIN2.SUB1
DOMAIN2.SUB2
USER A
USER B
USER D
(TEXT FILES OF DICTIONARY ENTRIES)
TRANSLATION PROCESSING MODULE
21
(TRANSLATION RECORDS)
(OUTPUT TEXT)
(INPUT TEXT)
DICTIONARY CONTROL MODULE
13
13a
DICTIONARY SELECTION
INFERENCING ENGINE
(CORE LANG. PAIR SELECTION)
(SUBLANG. DICT. SELECTION)
13b
DICTIONARY MAINTENANCE
USER MAINTENANCE
INPUT EDITOR
(SCRATCH ENTRY)
13b-1
(INTERACTION RECORDS)
13b-2
DMO MAINTENANCE
DICTIONARY MAINTENANCE UTILITY
DMO ASSISTANCE UTILITIES
(INPUT TEXT)
(HEADER DATA)
RECOGNITION MODULE
12
OCR
USER ID FILES
11
RECEIVING INTERFACE
INTERACTIVE MODE
(INPUT DATA)
(INTERACTION ON NEW WORDS)
A
B
30
OUTPUT MODULE

FIG. 4

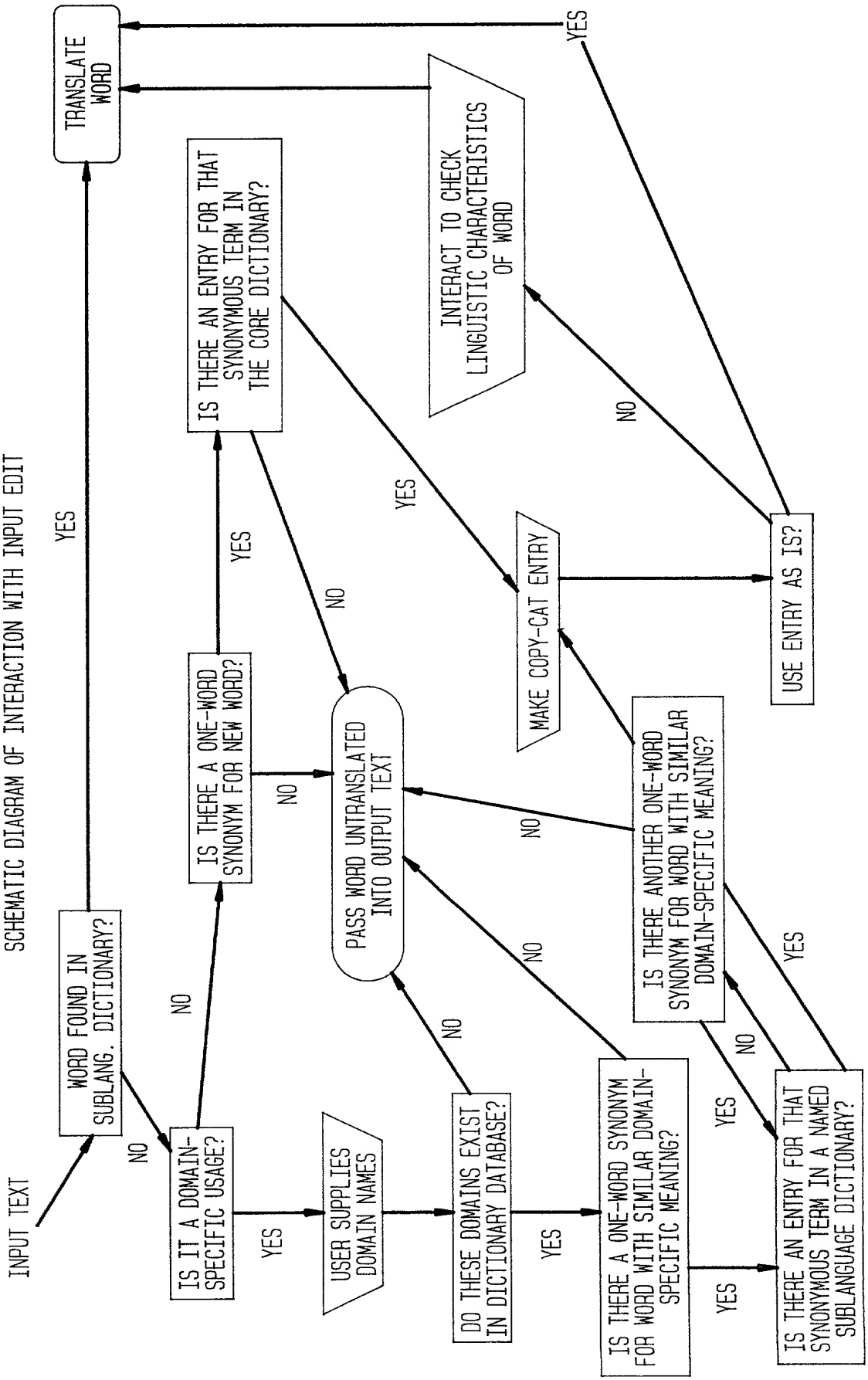
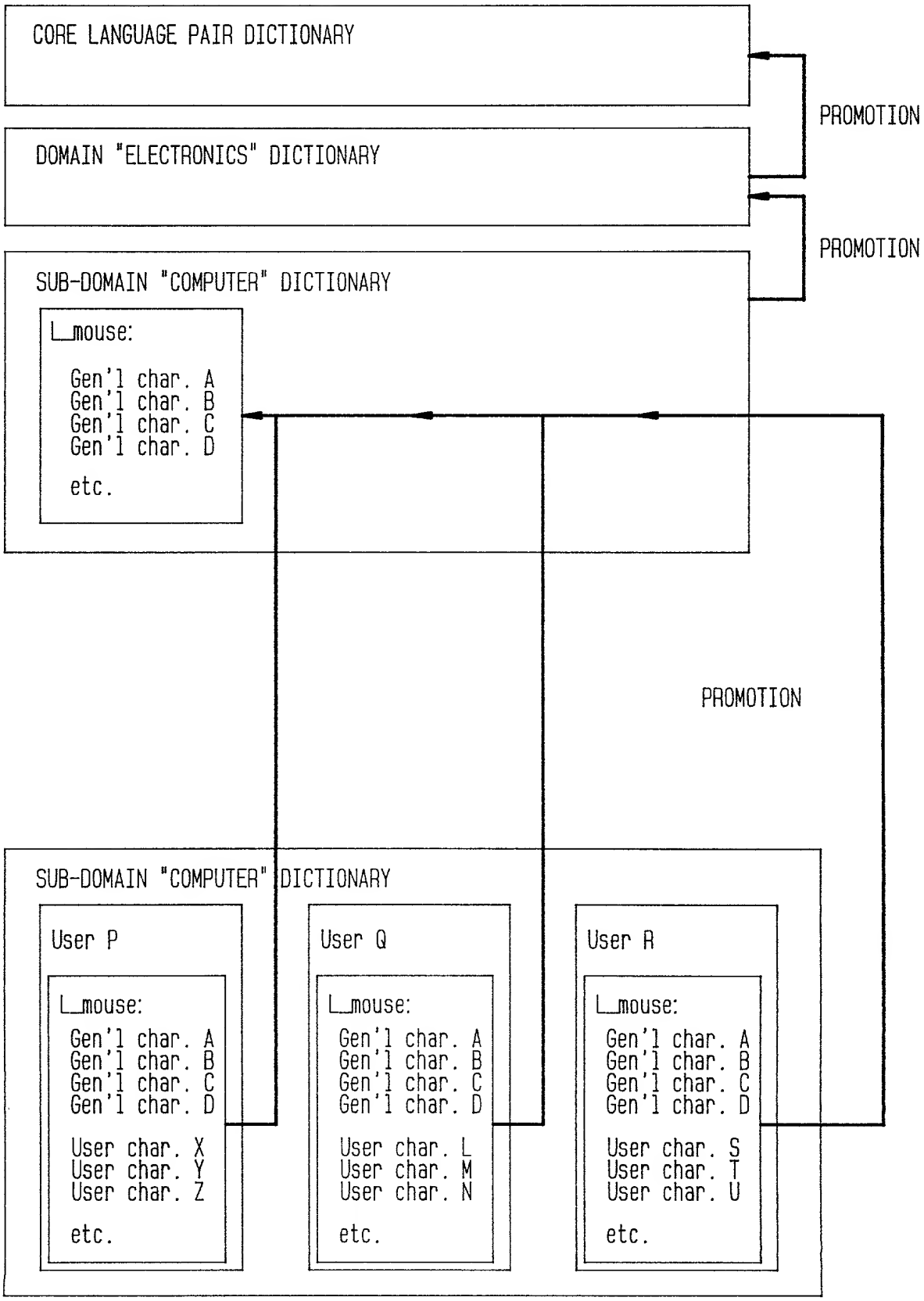


FIG. 5

SCHEMATIC DIAGRAM OF INTERACTION WITH INPUT EDIT

FIG. 6

COLLAPSING AND PROMOTION OF ENTRIES FROM SUBORDINATE TO SUBORDINATE DICTIONARIES



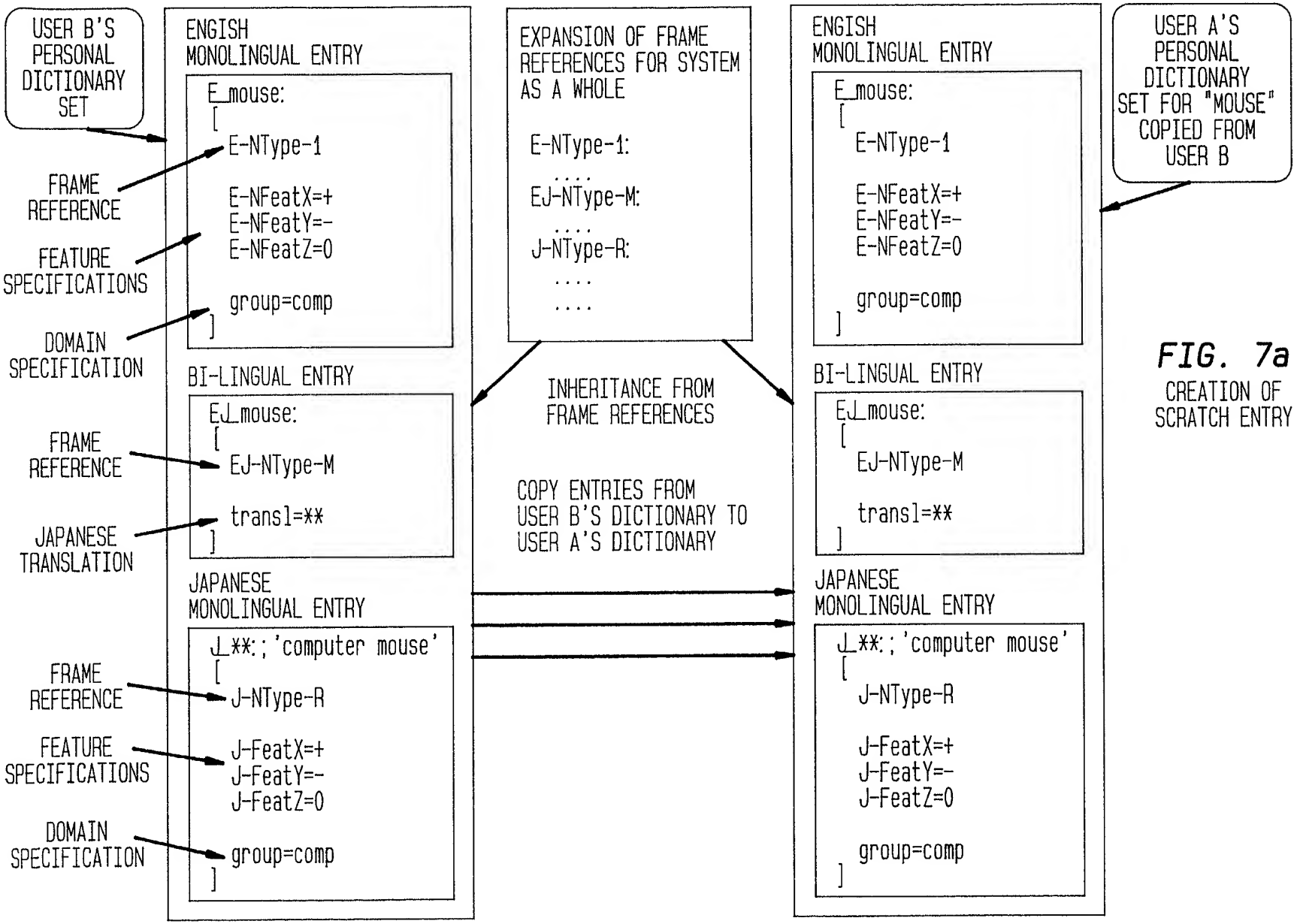


FIG. 7a
CREATION OF SCRATCH ENTRY

FIG. 7b

SCANNING PROCESS OF LEVELS OF DICTIONARIES BELOW AND ABOVE DOMAIN SUPPLIED TO INPUT EDITOR BY USER A

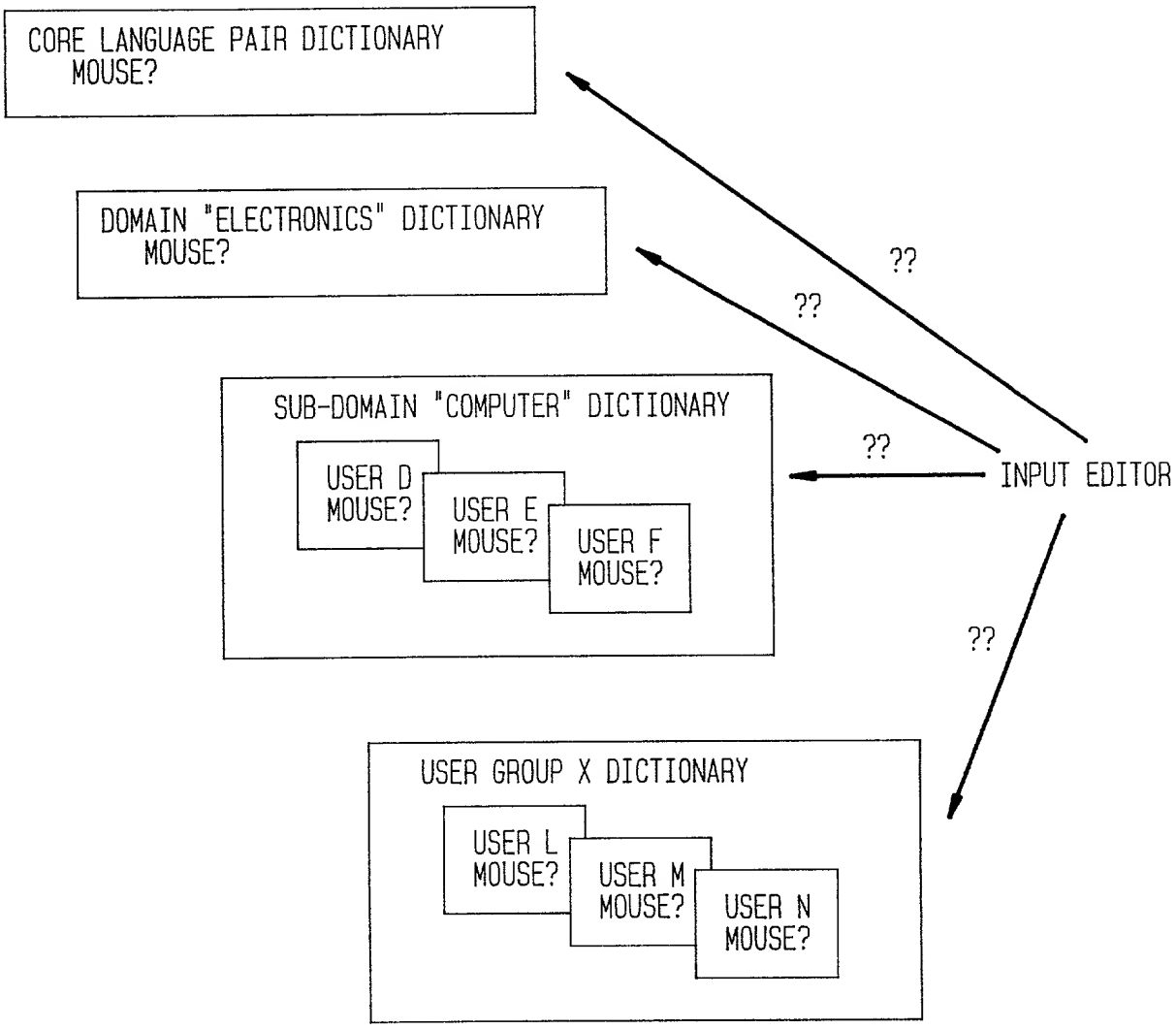
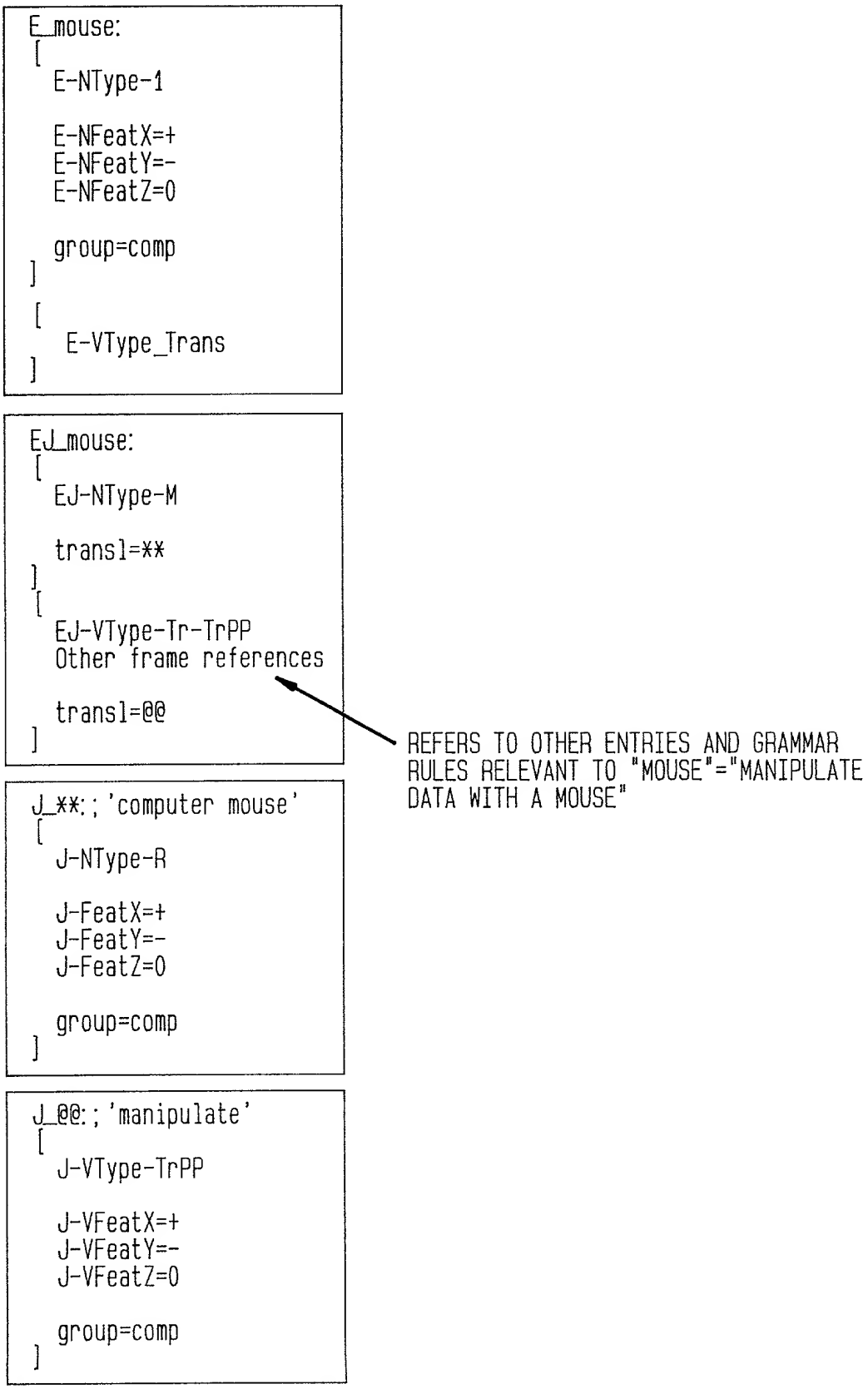


FIG. 7c

USER C's ENTRIES RELEVANT TO "MOUSE"



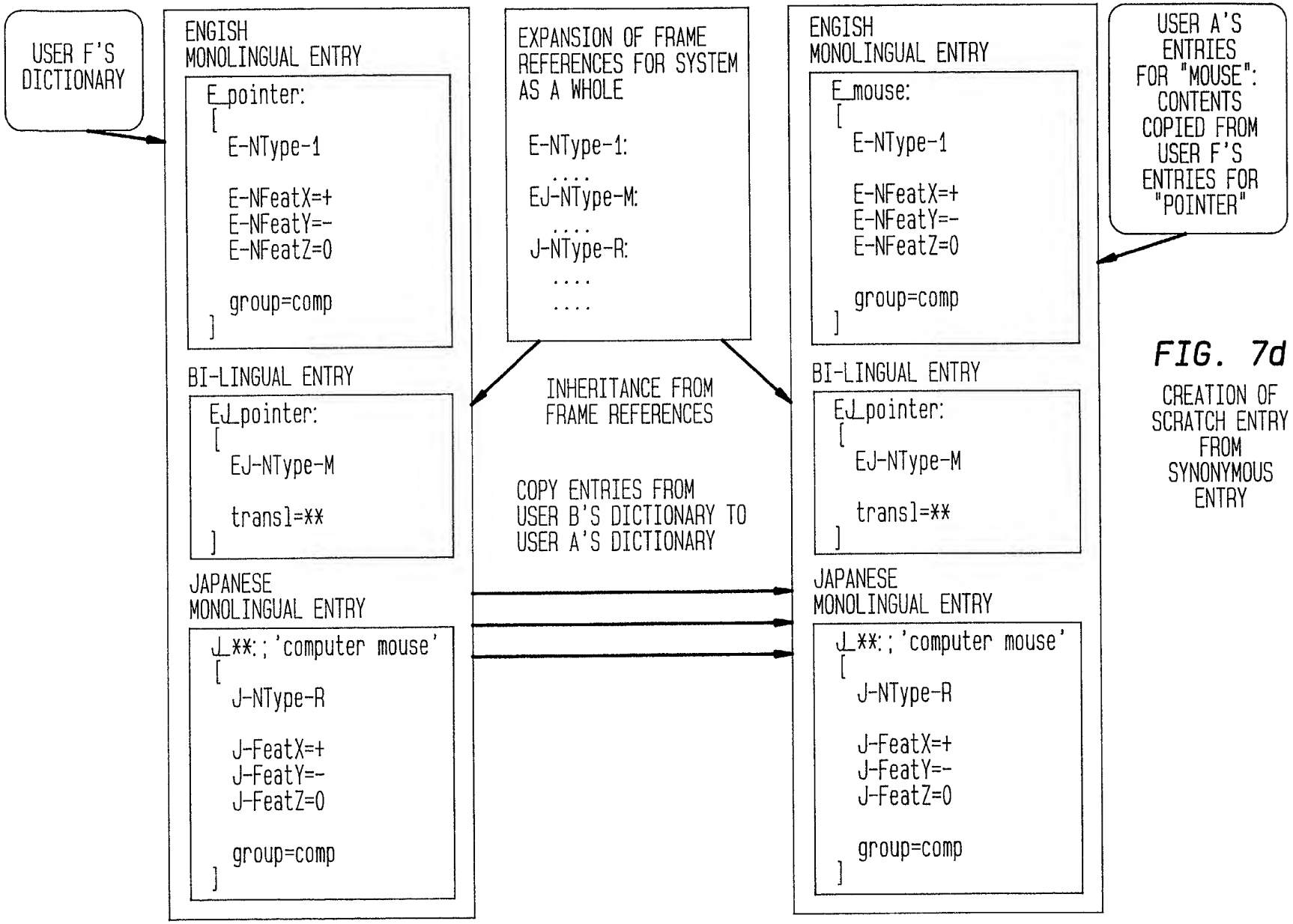


FIG. 7d

CREATION OF SCRATCH ENTRY FROM SYNONYMOUS ENTRY

PTO/SB/51 (12-97)
Approved for use through 9/30/00. OMB 0651-0033
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REISSUE APPLICATION DECLARATION BY THE INVENTOR

Docket Number (Optional): TLC-RE

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are stated below next to my name.
I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is described and claimed in patent number USP 5,535,120, granted July 9, 1996, and for which a reissue patent is sought on the invention entitled MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ID DATA TO SELECT DICTIONARIES,
the specification of which

☒ is attached hereto.

☐ was filed on _____ as reissue application number ___ / _____
and was amended on _____.
(If applicable)

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.
I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56.
I verily believe the original patent to be wholly or partly inoperative or invalid, for the reasons described below. (Check all boxes that apply.)

☐ by reason of a defective specification or drawing.

☒ by reason of the patentee claiming more or less than he had the right to claim in the patent.

☐ by reason of other errors.

At least one error upon which reissue is based is described as follows:

The issued claims refer to the selection of dictionaries for machine translation by user ID data, but do not refer to selection of source/target languages by user ID data or by text routing information, which was within the main concept of the invention. Thus, the issued claims are believed to be insufficient.

PTO/SB/51 (12-97)
Approved for use through 9/30/00. OMB 0651-0033
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

(REISSUE APPLICATION DECLARATION BY THE INVENTOR, page 2)

Docket Number (Optional): TLC-RE

All errors corrected in this reissue application arose without any deceptive intention on the part of the applicant. As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Name(s)	Registration Number
Leighton K. Chong	27,621

Correspondence Address: Direct all communications about the application to:

☐ Customer Number _____ → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm or Individual Name	Leighton K. Chong, Firm of Ostrager Chong Flaherty & Onofrio				
Address	841 Bishop Street, Suite 1200				
Address					
City	Honolulu	State	HI	ZIP	96813-3908
Country	U.S.A.				
Telephone	(808) 533-4300	Fax	(808) 531-7585		

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine and imprisonment, or both, under 18 U.S.C. 1001, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this declaration is directed.

Full name of sole or first inventor (given name, family name): Leighton K. CHONG

Inventor's signature: Leighton K. Chong

Residence: 133 Kaai Street, Honolulu, HI — Date: June 22, 1998

Post Office Address: (same) — Citizenship: U.S.A.

Full name of second joint inventor (given name, family name): Christine K. KAMPRATH

Inventor's signature: Christine K. Kamprath — Date: June 26, 1998

Residence: 100 Oakwood Park, Peoria, IL — Citizenship: U.S.A.

Post Office Address: (same)

Full name of third joint inventor (given name, family name):

Inventor's signature — Date

Residence — Citizenship

Post Office Address

☐ Additional joint inventors are named on separately numbered sheets attached hereto.

PTO/SB/09 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office; U S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) & 1.27(b))–INDEPENDENT INVENTOR	Docket Number (Optional) TLC-RE

Applicant, Patentee, or Identifier: L. CHONG, C. KAMPRATH

Application or Patent No.: USP 5,535,120 (Reissue)

Filed or Issued: July 9, 1998

Title: MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ID DATA TO SELECT DICTIONARIES

As a below named inventor, I hereby state that I qualify as an independent inventor as defined in 37 CFR 1.9(c) for purposes of paying reduced fees to the Patent and Trademark Office described in:

☐ the specification filed herewith with title as listed above.

☐ the application identified above.

☒ the patent identified above.

I have not assigned, granted, conveyed, or licensed, and am under no obligation under contract or law to assign, grant, convey, or license, any rights in the invention to any person who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person had made the invention, or to any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

☐ No such person, concern, or organization exists.

☒ Each such person, concern, or organization is listed below.

TRANS-LINK INTERNATIONAL CORP., a small entity,

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

Leighton K. CHONG	Christine K. KAMPRATH	
NAME OF INVENTOR	NAME OF INVENTOR	NAME OF INVENTOR
Leighton K Chong	Christine K Kamprath	
Signature of inventor	Signature of inventor	Signature of inventor
June 22, 1998	June 26, 1998	
Date	Date	Date

Burden Hour Statement: This form is estimated to take 0 2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/10 (12-97)
Approved for use through 9/30/00. OMB 0651-0031
Patent and Trademark Office: U S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) & 1.27(c))--SMALL BUSINESS CONCERN

Docket Number (Optional): TLC-RE

Applicant, Patentee, or Identifier: L. CHONG, C. KAMPRATH
Application or Patent No.: USP 5,535,120 (Reissue)
Filed or Issued: July 9, 1998
Title: MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ID DATA TO SELECT DICTIONARIES

I hereby state that I am
☐ the owner of the small business concern identified below:
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN TRANS-LINK INTERNATIONAL CORP.

ADDRESS OF SMALL BUSINESS CONCERN 133 Kaai Street
Honolulu, HI 96821

I hereby state that the above identified small business concern qualifies as a small business concern as defined in 13 CFR Part 121 for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby state that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention described in:

☐ the specification filed herewith with title as listed above.
☐ the application identified above.
☒ the patent identified above.

If the rights held by the above identified small business concern are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d), or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:
☒ no such person, concern, or organization exists.
☐ each such person, concern, or organization is listed below.

Separate statements are required from each named person, concern or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING Leighton K. CHONG

TITLE OF PERSON IF OTHER THAN OWNER President

ADDRESS OF PERSON SIGNING 133 Kaai Street, Honolulu, HI 96821

SIGNATURE Leighton K. Chong DATE June 22, 1998

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/53 (12-97)
Approved for use through 9/30/00. OMB 0651-0033
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REISSUE APPLICATION BY THE INVENTOR, OFFER TO SURRENDER PATENT

Docket Number (Optional)

This is part of the application for a reissue patent based on the original patent identified below.

Name of Patentee(s)
Leighton K. CHONG, Christine K. KAMPRATH

Patent Number	Date Patent Issued
USP 5,535,120	July 9, 1998

Title of Invention
MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ...

I am the inventor of the original patent.

I offer to surrender the original patent.

1. [XX] Filed herein is a certificate under 37 CFR 3.73(b).

2. [ ] Ownership of the patent is in the inventor(s), and no assignment of the patent has been made.

One of boxes 1 or 2 above must be checked.

The written consent of all assignees owning an undivided interest in the original patent is included in this application for reissue.

Signature	Date
Leighton K Chong	June 22, 1998

Typed or printed name
Leighton K. CHONG

The assignee owning an undivided interest in said original patent is TRANS-LINK INT'L CORP., and the assignee consents to the accompanying application for reissue.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application, any patent issued thereon, or any patent to which this declaration is directed.

Name of assignee
TRANS-LINK INTERNATIONAL CORP.

Signature of person signing for assignee	Date
Leighton K Chong	June 22, 1998

Typed or printed name and title of person signing for assignee
Leighton K. Chong, President

Burden Hour Statement: This form is estimated to take 0.1 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/53 (12-97)
Approved for use through 9/30/00. OMB 0651-0033
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

REISSUE APPLICATION BY THE INVENTOR, OFFER TO SURRENDER PATENT	Docket Number (Optional)

This is part of the application for a reissue patent based on the original patent identified below.

Name of Patentee(s)
Leighton K. CHONG, Christine K. KAMPRATH

Patent Number	Date Patent Issued
USP 5,535,120	July 9, 1998

Title of Invention
MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM USING USER ...

I am the inventor of the original patent.

I offer to surrender the original patent.

1. [XX] Filed herein is a certificate under 37 CFR 3.73(b).

2. [ ] Ownership of the patent is in the inventor(s), and no assignment of the patent has been made.

One of boxes 1 or 2 above must be checked.

The written consent of all assignees owning an undivided interest in the original patent is included in this application for reissue.

Signature	Date
Christine K. Kamprath	June 26, 1998

Typed or printed name
Christine K. KAMPRATH

The assignee owning an undivided interest in said original patent is TRANS-LINK INT'L CORP., and the assignee consents to the accompanying application for reissue.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application, any patent issued thereon, or any patent to which this declaration is directed.

Name of assignee
TRANS-LINK INTERNATIONAL CORP.

Signature of person signing for assignee	Date
Leighton K. Chong	June 22, 1998

Typed or printed name and title of person signing for assignee
Leighton K. Chong, President

Burden Hour Statement: This form is estimated to take 0.1 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application For Reissue Of	Atty. #: TLC-RE
CHONG	Examiner:
For U.S. Patent 5,535,120 iss. July 9, 1996	Group No.:
Filed: (Concurrently Herewith)	
Title: MACHINE TRANSLATION AND TELECOMMUNICATIONS SYSTEM	

ORDER FOR A TITLE REPORT

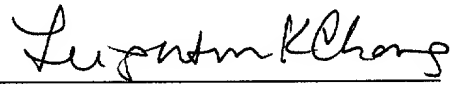
Assistant Commissioner for Patents
U.S. Patent & Trademark Office
Washington, D.C. 20231

Sir:

With the filing of the above-identified application for reissue, applicant requests that a title report for the patent be ordered and placed in the file, pursuant to 37 C.F.R. 1.171.

Authorization is hereby given to charge our Deposit Account No. 15-0699 for the title report fee.

Respectfully submitted,
ATTORNEYS FOR APPLICANT



Leighton K. Chong
OSTRAGER CHONG FLAHERTY & ONOFRIO
841 Bishop Street, Suite 1200
Honolulu, HI 96813-3908
Tel: (808) 533-4300 (H.S.T.)

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division

Application deficiencies found during scanning:

1. Application papers are not suitable for scanning and are not in compliance with 37 CFR 1.52 because:
   - ☐ All sheets must be either A4 (21 cm x 29.7 cm) or 8-1/2"x 11".
     Pages _____ do not meet these requirements.
   - ☐ Papers are not ☐ flexible, ☐ strong, ☐ smooth, ☐ non-shiny, ☐ durable, and ☐ white.
   - ☐ Papers are not ☐ typewritten or mechanically printed ☐ in permanent ink ☐ on one side.
   - ☐ Papers contain improper margins. Each sheet must have a left margin of at least 2.5 cm (1") and top, bottom and right margins of at least 2.0 cm (3/4").
   - ☐ Papers contain hand lettering.

2. Drawings are not in compliance and were not scanned because:
   - ☐ The drawings or copy of drawings are not suitable for electronic reproduction.
   - ☐ All drawings sheets are not either A4 (21 cm x 29.7 cm) or 8-1/2" x 11".
   - ☐ Each sheet must include a top and left margin of at least 2.5 cm (1"), a right margin of at least 1.5 cm (9/16") and a bottom margin of at least 1.0 cm (3/8").

3. Page(s)_____ are not of sufficient ☐ clarity, ☐ contrast and ☐ quality for electronic reproduction.

4. Page(s)_____ are missing.

5. OTHER: No Transmittal